THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXX October, 1939 Number 7

AGE AND SEX PATTERNS OF SOCIAL ATTITUDES

A. C. ROSANDER[1]

United States Bureau of Labor Statistics

The Maryland youth survey, the final report of which is now available,[2] was concerned not only with the social and economic characteristics of youth sixteen to twenty-five years of age,[3] but with their attitudes toward important social and economic questions.

One of the most significant aspects of these data on attitudes is how the latter changed as the youth increased in age. The data are extensive, since for each of the nine age-groups between thirteen hundred and sixteen hundred youth were interviewed, about equally divided between the sexes.[4] While age is the major factor under consideration, the effect of sex will be included in all but a few questions. How the attitudes of youth from different types of home and educational backgrounds change as they grow older is a question of basic significance for which data are available, but a treatment of it is beyond the scope of this discussion.

Some of the attitudes which we shall discuss will represent overt behavior, while others will represent what are commonly called beliefs or opinions. In the former category are included drinking intoxicating beverages to any degree, voting at the last opportunity, and attending church. In the latter group are attitudes toward war, child labor, government relief for the unemployed, government regulation of minimum wages and maximum hours in business and industry, the

[1] Formerly statistician and research associate, Maryland Youth Survey.

[2] The report entitled "Youth Tell Their Story" by Howard Bell is two hundred seventy-three pages in length, contains one hundred tables, twenty-three graphs, and thirty-nine pictograms, and is sold by the American Youth Commission, 744 Jackson Place, Washington, D. C.

[3] In this study age is that at last birthday.

[4] The writer is indebted to Dr. Homer P. Rainey, former director of the American Youth Commission, for permission to use hitherto unpublished data.

election of capable men to public office, and sex education in the schools. The attitude of all youth toward the economic value of his schooling, and the attitude of employed youth toward the wages which he was receiving at the time of the interview, which was during the Summer or Fall of 1936, were also investigated.

We shall present our findings under two classifications: (*a*) Those attitudes which indicated approximately *constant* proportions of acceptance (or rejection) at the various ages, and (*b*) those attitudes which indicated *changing* proportions of acceptance (or rejection) at the various ages. We shall also discuss certain implications of the findings and consider some hypotheses which these data seem to suggest.

ATTITUDES SHOWING CONSTANT PROPORTIONS BY AGE

By constant proportions we mean that the line $y = a$ is an adequate fit to the nine percentages each being the proportion of the sampled youth at a given age with the specified attitude.

One way of determining whether $y = a$ is applicable to the data is to examine the graph showing the trend of the proportions by age for each specific attitude. A more rigorous method is to test whether the variations in the proportions by age could have arisen from fluctuations in samples drawn from the same attitude universe. A third method is that of determining whether the slope of a straight line fitted to the data by the method of least squares or some other method, is significantly different from zero.

Due to the relative high stability of the proportions the use of the first two methods was deemed sufficient. It was not thought advisable to apply the method of least squares since the age range is narrow and any generalization or extrapolation is therefore likely to be dangerous.

Certain specific problems arise in determining whether or not a set of proportions could have been drawn from the same population. In the first place they may vary over a narrow, or over a wide, range and still show no age trend. They may rise (or fall) slowly over the various ages to the extent of only three or four per cent in absolute change, which may indicate such a moderate shift in attitudes as to require data from a much wider age range in order to get a very noticeable change. It is obvious that a nine-year section of a given attitude out of a fifty-year range may appear to be constant even though other sections may be highly curvilinear. Again, the first and last ages of our age group may be on opposite sides of the constant trend of the

intervening ages and, hence, give the impression of a trend. Without additional age data it is difficult to conclude whether these end differences are real or not. We also have some instances of a sharp break in proportions at a specific age; whether this apparent discontinuity is characteristic of certain attitudes at certain ages can be established only through further investigation.

In all tables the number of youth given represent the size of the sample, *not* the number of youth giving a specified response, and are designated by the heading "percentage base." If the number of youth giving a specified response is desired, it can be obtained by multiplying the percentage base by the per cent giving the response in question.

In order to test whether or not the proportions of youth at the different ages who accepted (or rejected) a given attitude could have arisen from fluctuations in samples drawn from the same attitude universe, we assumed that the percentage for the entire sample is the best estimate of the proportion in the universe, that the percentages are normally distributed, and tested whether any of the specific age samples deviated from this percentage more than would be expected on the basis of random Bernoullian sampling. For this purpose we employed the usual formula for the variance of a sampled percentage: $\sigma_p^2 = \frac{p(1-p)}{N}$ where p is the true proportion in the universe for which we have used the grand percentage as the best estimate, and N is the number of youth in any given age or age-sex sample under consideration.

KIND AND LEVEL OF GOVERNMENT RELIEF

The youth were questioned with regard to the kind and level of relief, if any, that the government should provide the needy unemployed. By health and decency level was meant something more than that necessary for food, housing, and clothing such as money to maintain health and provide for an occasional movie. The data indicate that among male youth, eight per cent were opposed to any form of government relief, eighty-nine per cent favored government work relief, while eighty-one per cent believed relief should be on a health and decency level. The corresponding percentages among female youth were nine, eighty-eight, and eighty-three, respectively.

Tests applied to the percentages of those favoring work relief, assuming the grand percentage to be the true value, show that for the male youth the greatest positive deviation is 0.023 (for age eighteen)

while the greatest negative deviation is 0.025 (for age twenty-two). The standard error of the assumed true proportion of 0.890 for samples giving these deviations is 0.012; therefore, both these deviations represent close to two standard errors from the mean. Since the other seven samples deviate less than these two, this is evidence that all these samples could have been drawn from the same attitude universe. A similar situation exists relative to the female youth where the largest deviations are 0.013 and 0.017 measured from the grand percentage of 0.882; the larger deviation is approximately one and five-tenths standard errors from the mean.

Similar tests applied to the percentages favoring a health and decency level reveal that σ_p for the male and female samples which deviate most from the mean is 0.014; since the maximum deviations for the male youth are 0.025 and 0.026, and for the female youth are 0.020 and 0.024; these are well within two standard errors of the mean.

The most important point relative to the relief question is the high and fairly constant proportions in favor of work relief on a decency level for each sex for each age; neither age nor sex variation is as large for these questions as it is for some of the others to be discussed. While a higher percentage of female youth at every age but two favored relief on a decency level, the difference is small, averaging approximately two per cent.

MINIMUM WAGES AND MAXIMUM HOURS

The responses to the question "In reference to the fixing of minimum wages and maximum hours, to what degree should the government participate?" showed that fifty-four per cent of each sex was in favor of such a measure applied to all business and industry. The proportion of youth who indicated the four major responses to the question, and the number of youth on which these percentages are based, are shown in Table I. Since the standard deviation of this percentage is approximately 0.018 all deviations for specific age-sex samples are within two standard deviations of the true value of their sex group, with the possible exceptions of the twenty-three- and twenty-four-year-old male groups and the seventeen-year-old female group; the former are slightly beyond two standard errors while the latter is below three standard errors from the assumed true value. Ignoring these exceptions, the favorable attitude of these youth relative to government regulation of minimum wages and maximum hours is almost as prevalent in one age-sex group as in another.

An eight per cent absolute range in variability of proportions occurs in each sex group; this is about twice as great as that observed for the relief questions, and at a lower magnitude. In addition it appears that the older male youth (twenty-two to twenty-four years), probably because a greater proportion is employed, to the extent of four per cent are more in favor of this type of government regulation than are the younger male youth.

It is significant to note that more than half the youth in all the groups of voting age were in favor of a more far-reaching law than the one passed by Congress in June, 1938.

TABLE I.—ATTITUDE TOWARD GOVERNMENT REGULATION OF MINIMUM WAGES AND MAXIMUM HOURS IN BUSINESS AND INDUSTRY—MARYLAND, 1936

Age	Per cent with specified attitude[1]								Percentage base	
	Yes, in all business		Yes, in some business		No regulation at all		No opinion			
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
16	51.7	55.4	23.7	18.0	16.0	13.0	8.6	13.6	776	724
17	53.3	48.8	23.4	23.6	16.5	13.1	6.8	14.5	722	686
18	56.0	55.5	22.1	19.6	15.7	14.3	6.2	10.6	712	761
19	50.4	50.8	21.4	20.8	17.0	11.3	11.2	17.1	863	771
20	53.8	57.2	22.6	16.8	15.7	12.1	7.9	13.9	810	820
21	50.8	57.2	23.0	14.6	18.1	14.5	8.1	13.7	817	785
22	56.3	54.8	20.7	19.7	15.0	13.1	8.0	12.4	695	681
23	58.2	56.0	17.2	18.9	14.7	12.2	9.9	12.9	639	588
24	57.7	54.2	18.1	18.9	13.4	12.9	10.8	14.0	789	731
Total..	54.1	54.5	21.4	18.9	15.8	13.0	8.7	13.6	6823	6547

[1] Within each age group the four percentages for each sex add to 100.0.

CHILD LABOR

By child labor is meant the gainful employment of children under sixteen years of age. The question used by the interviewers was as follows: "Under what circumstances, if any, should government permit the gainful labor of children under sixteen years of age?" The total number of male and female youth answering this question in each age group, and the proportions indicating the responses specified in the column headings are shown in Table II.

Instead of testing the nine age samples in each sex group it is sufficient to compute the lowest and the highest values of σ_p for each sex group. For the male youth we find that the lowest value of σ_p to be 0.017 (for age nineteen) and the highest value to be 0.020 (for age twenty-three); this means that no age sample deviates as much as $3\sigma_p$ from .471, while seven out of the nine samples deviate less than $2\sigma_p$. Furthermore, the deviations from .471 give no evidence of a trend in the observations and lead to a standard error of the sample residuals of 2.6, with the largest deviation being one and five-tenths times as great as this value. This is evidence that these age samples might well have been drawn from the same attitude universe in which p is .471 (see Table II).

For the female youth the lowest value of the standard error of p is 0.018 (for age twenty), while the highest value is 0.021 (for age twenty-three), with no single proportion as much as two standard errors removed from .491. The standard deviation of the sample residuals is 2.4 with the largest deviation being not quite one and five-tenths times as great as this value. This group of youth, like the male youth, could have been drawn from the same attitude universe.

TABLE II.—ATTITUDE TOWARD THE GAINFUL EMPLOYMENT OF CHILDREN UNDER SIXTEEN YEARS OF AGE—MARYLAND, 1936

Age	Per cent of each age-sex group[1]								Percentage base	
	No child labor		If family needs financial aid		Under other permissive circumstances		Under any circumstance			
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
16	44.2	52.0	37.7	37.3	12.7	7.5	5.4	3.2	764	713
17	47.8	46.5	35.4	39.9	12.8	9.6	4.0	4.0	718	676
18	49.6	52.6	35.0	34.7	11.6	9.9	3.8	2.8	708	755
19	43.1	47.6	40.6	39.3	11.4	10.4	4.9	2.7	854	760
20	51.0	48.1	32.7	38.1	12.3	11.0	4.0	2.8	798	804
21	49.5	49.0	34.3	36.6	11.9	11.5	4.3	2.9	810	787
22	47.4	45.8	37.2	38.0	12.0	13.3	3.4	2.9	685	677
23	47.0	47.7	32.9	35.4	14.9	14.8	5.2	2.1	630	579
24	44.5	51.7	34.9	31.7	17.6	13.9	3.0	2.7	774	725
Total..	47.1	49.1	35.7	36.8	13.0	11.2	4.2	2.9	6741	6476

[1] Within each age group the four percentages for each sex add to 100.0.

Applying the standard error of a percentage formula to each of the sex groups for all ages combined, we find that the difference of two per cent could hardly have arisen from random sampling, and hence that a small but real difference between the sexes probably exists.

While the fluctuation of percentages by age covers an eight per cent band, and a slightly higher proportion of the female youth is opposed to child labor, still no age bias exists among the female youth, although there probably is some bias in the male youth from twenty to twenty-four years of age. This may be due to the fact that a greater proportion of farm workers was included in ages twenty-three and twenty-four than in the younger age groups.

SEX EDUCATION IN SCHOOLS

The question as to whether sex information should be offered in the school was answered in the affirmative by approximately three-fourths of all the youth. Several points are to be observed from the data given in Table III.

The first two ages for the male youth are on a slightly lower level than the average of the remaining seven ages. For the female youth the sixteen-year-old group is definitely lower than the average and gives evidence of an upward trend, while the twenty-three- and twenty-four-year-old groups tend to be somewhat higher than the other groups. The difference of eleven per cent between the sixteen-year-olds and the twenty-three-year-olds is real, and cannot be explained by sampling fluctuations. However, between the ages of seventeen and twenty-two, there is no trend, but only an attitude plateau.

Another characteristic is the widening gap between the sexes beyond twenty years of age which may be explained to some degree by the larger proportion of married female youth than married male youth in these upper age groups, who as a result of marriage may have become conscious of the need for definite sex education.

OPINION TOWARD WAR

Youth were asked their opinion toward war and, on the basis of the responses, were classified by the interviewers according to their principal reaction. One of these categories was the statement or its equivalent that "war is needless and preventable." With regard to this specific response the data in Table III show some very interesting relationships.

The approximately constant difference of twelve per cent between the two sexes shows that sex and not age is the significant factor in this case. Even so, more than half of the male youth expressed this attitude toward war. Whether the downward trend observable in both sexes beyond the age of twenty-one continues into the late twenties and thirties is an interesting but unanswered conjecture.

TABLE III.—ATTITUDES TOWARD HOW CANDIDATES FOR PUBLIC OFFICE ARE ELECTED, TOWARD WAR, AND TOWARD SEX EDUCATION IN THE SCHOOLS—MARYLAND, 1936

Age	Per cent and number of each age-sex group with specified attitude									
	"Candidates are not elected by merit"		"War is needless and preventable"				"Yes, have sex education in the schools"			
	Per cent	Percentage base	Per cent		Percentage base		Per cent		Percentage base	
			Male	Female	Male	Female	Male	Female	Male	Female
16	28.5	1514	52.9	67.0	781	730	71.7	70.6	782	732
17	31.5	1423	57.0	67.6	723	691	69.9	73.7	727	696
18	30.6	1483	57.0	66.1	711	762	76.7	78.3	717	766
19	29.6	1656	54.3	67.7	867	780	72.9	73.4	873	783
20	31.4	1643	54.9	66.8	810	826	77.6	77.7	814	829
21	34.1	1620	56.6	68.8	818	794	73.3	75.5	824	796
22	31.3	1401	55.4	67.8	702	695	71.3	75.3	704	697
23	32.3	1242	51.9	65.4	642	595	75.7	81.9	646	596
24	31.7	1546	52.2	64.5	797	743	74.3	78.2	802	744
Total..	31.1	13528	54.7	66.9	6851	6616	73.6	76.0	6889	6639

The narrow range of proportions for the female youth in contrast with the greater fluctuations in those of male youth, shows that the former, regardless of age, represent a more homogeneous attitude group than the latter. It is interesting to compare this question which does not show an age trend with the other war question in which youth were asked what they would *do* in case war was declared, in which an age variation was discovered. (See Table VI.)

CANDIDATES FOR PUBLIC OFFICE

The last of the questions which illustrated a constant trend of proportions by age is that in which youth stated his attitude as to how candidates were elected to public office. We report here that group of youth who thought that candidates were elected not on the basis of personal merit to fill the office competently, but by such means as "pull," money, false promises, political machines. The various proportions for each age, both sexes combined, are shown in Table III. With the exception of the sixteen-year-olds who are slightly low (two and six-tenths per cent below the mean), and the twenty-one-year-olds who are slightly high (three per cent above the mean) no age group deviates very much from thirty-one per cent.

ATTITUDE TRENDS OF THE SEXES COMBINED

When for each of the several specific attitudes already discussed the percentages for both sexes are combined into a single figure, this proportion is approximately halfway between the two percentages since the sample sizes of the two sexes are not widely different.

The obvious effect of making this combination is to reduce the within-age variations. The only attitude in which any trend is observable is that which concerns sex education in schools, and this, no doubt, reflects the difference between male and female youth at the upper ages, a difference already commented upon.

The seven specific attitudes covered involve social, economic, and political questions; and yet for youth of these nine ages the proportions with the given attitude show relatively small variations about a mean value. A discussion of the implications of these data will be postponed until we have presented the questions which show a definite age trend, either in terms of increasing or decreasing proportions.

ATTITUDES SHOWING CHANGING PROPORTIONS BY AGE

Drinking.—Drinking as defined in the survey meant taking intoxicating beverages to whatever degree however slight; no attempt was made to distinguish drinking on special occasions, moderate drinking, or extreme drinking. Data were obtained not only with regard to those who drank and those who did not, but also the division of the latter group into those who were not opposed to drinking in contrast to those who were opposed. The proportion of each sex at each age

group who answered in one of these three specified ways is shown in Table IV.

TABLE IV.—ATTITUDE OF YOUTH TOWARD DRINKING INTOXICATING BEVERAGES (PER CENT OF EACH AGE-SEX GROUP)[1] MARYLAND—1936

Age	Drinks[2]		Does not drink				Percentage base	
			Not opposed		Opposed			
	Male, per cent	Female, per cent	Male, per cent	Female, per cent	Male, per cent	Female, per cent	Male	Female
16	37.6	26.1	41.3	41.7	21.1	32.2	777	723
17	44.6	27.7	34.9	39.5	20.5	32.8	716	693
18	53.3	39.0	28.2	33.6	18.5	27.4	709	758
19	61.5	43.4	23.1	28.5	15.4	28.1	864	771
20	63.8	51.4	21.9	25.4	14.3	23.2	805	820
21	70.1	53.5	19.6	26.6	10.3	19.9	818	779
22	71.6	54.1	17.7	26.6	10.7	19.3	700	684
23	68.3	55.2	20.6	26.1	11.1	18.7	637	587
24	74.8	52.6	17.5	29.8	7.7	17.6	795	732
Total.....	60.7	44.7	24.9	30.8	14.4	24.5	6821	6547

[1] The three sex groups within each age add horizontally to 100.0 per cent.
[2] This means drinking to any extent or degree however slight.

The trend of the proportions indicates that opposition to drinking as here defined decreases as age increases, among female youth as well as among male youth. The former group never becomes as predominately tolerant as the latter group, but the general trend is the same.

Since the drinking group among female youth make a net absolute gain of twenty-six and five-tenths per cent between the ages of sixteen and twenty-four, this means that the group opposed to drinking lost members to the drinking group as age increased. The same trend is also true of the percentages for the male youth. This inference as well as the others made elsewhere in this discussion obviously are based upon the assumption that an age group is representative of any preceding age group when the latter reaches the age in question.

Most of the absolute gain in drinking among male youth occurred between the ages of sixteen and twenty-one; among female youth

between seventeen and twenty. With the two sexes combined, thirty-two per cent of the sixteen-year-olds stated they drank, while sixty-four per cent of the twenty-four-year-olds gave a similar answer. The fact that both curves of percentages are flattening out after the age of twenty is reached suggests that this type of behavior is a development of the late teens; it is possible, of course, that additional gains are made during ages beyond twenty-four.

Voting.—Information relative to whether or not the youth eligible to vote had voted at the last opportunity, and the reasons why the non-voters did not vote were obtained from forty-one hundred eighty-four youth, twenty-two, twenty-three, and twenty-four years of age. In most cases the last opportunity was a Spring or Fall election prior to the presidential election of 1936. Of these youth thirty-seven per cent of the twenty-two-year-olds, sixty per cent of the twenty-three-year-olds, and sixty-seven per cent of the twenty-four-year-olds had voted at the last opportunity. Of the entire group fifty-five per cent voted while forty-five per cent did not.

Of those who did not vote forty-four per cent gave lack of interest as the reason. This class increased from thirty-five per cent for the twenty-two-year-olds, to fifty-one per cent for the twenty-three-year-olds, and to fifty-two per cent for the oldest group. On the other hand, twenty-four per cent of the youth gave no reason at all for not voting, but this class decreased from thirty-five per cent to sixteen per cent to eleven per cent for the three age groups, respectively.

Church Attendance.—Youth were asked to indicate how often during the preceding twelve months they had attended church. These responses were classified into four groups: (1) Those who did not attend once, (2) those who attended on holidays only, (3) those who attended on the average once a month, and (4) those who attended on the average once per week. The proportions at each age for each sex that fell into each of these groups are shown in Table V.

The most marked change—and this appears to be true for each sex—is the steady decrease in the percentage of youth attending church once a week, and the increase in the proportions attending church less often or never. At age twenty-four male youth are almost equally divided among the four classes; while at this same age female youth show a leveling-off tendency relative to the early ages, but it is not so marked as that for the male youth. For both male and female youth about half of the absolute loss in the percentage of those going to church weekly took place between the ages of sixteen and nineteen.

TABLE V.—FREQUENCY OF CHURCH ATTENDANCE—MARYLAND, 1936

Age	Per cent of each age-sex group[1]								Percentage base	
	Never		Holidays only		Once a month		Once a week			
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
16	10.7	8.6	12.9	11.8	16.1	14.1	60.3	65.5	782	730
17	12.7	11.2	14.0	11.4	22.3	20.0	51.0	57.4	726	695
18	14.8	10.3	16.6	14.8	19.3	18.7	49.3	56.2	716	765
19	17.6	13.8	17.9	15.6	22.1	23.5	42.4	47.1	871	783
20	17.7	13.9	20.7	18.0	24.2	22.7	37.4	45.4	811	828
21	18.0	18.4	22.1	18.9	25.2	22.0	34.7	40.7	822	793
22	21.2	16.1	22.2	19.8	23.6	23.1	33.0	41.0	703	696
23	20.3	15.3	23.7	18.8	24.0	24.1	32.0	41.8	645	596
24	22.0	17.9	26.3	21.4	25.1	22.9	26.6	37.8	801	742
Total..	17.2	14.0	19.6	16.7	22.5	21.2	40.7	48.1	6877	6628

[1] Within each age group the four percentages for each sex add to 100.0.

War Action.—Youth were asked not only to express their opinions toward war, but also to state what action they believed they would take in case war were declared by this country. The age and sex proportions for each category of response are given in Table VI.

That both age and sex are factors influencing expressed war action is clearly evident from the data. Not only is there a steady decline in proportions of "volunteers" for each sex, but at every age the sex difference is pronounced, the female youth having the smaller proportions of war enthusiasts. The opposite position of refusing to go, or trying not to go, shows the reverse trend, the proportions by age increasing for each sex, with the female youth having the larger proportions of refusals.

With regard to the attitude "go if drafted" some, but not much, age difference is observable; sex is apparently a factor, with the female youth somewhat less inclined toward this view. Obviously, there is a correlation between proportions of responses, since if they are high for one response they must be low for other responses.

The diverging trend of the sex lines as age increase for both volunteers and refusals, indicates that as age increases female youth become more rapidly anti-war than do the male youth.

TABLE VI.—WAR ACTION YOUTH[1] INDICATED HE WOULD TAKE IN CASE WAR WERE DECLARED[1]—MARYLAND, 1936

Age	Per cent in each age-sex group[4]										Percentage base	
	Volunteer		Go if drafted[2]		Refuse to go[3]		Other action		Does not know			
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
16	46.5	34.5	38.3	40.8	10.3	14.8	1.2	1.6	3.7	8.3	780	731
17	39.6	29.5	46.7	44.9	8.9	17.9	1.5	0.9	3.3	6.8	723	692
18	37.4	29.7	47.8	41.4	9.9	19.6	0.6	2.0	4.3	7.3	715	765
19	36.7	28.2	47.9	42.0	10.0	18.4	0.8	1.9	4.6	9.5	869	781
20	33.0	23.5	47.9	45.6	13.4	20.1	1.9	1.2	3.8	9.6	812	826
21	34.4	21.4	46.7	42.3	12.4	23.8	1.1	2.6	5.4	9.9	815	791
22	31.1	20.6	50.5	41.6	13.8	24.3	1.2	2.9	3.4	10.6	701	690
23	32.3	18.2	49.5	47.9	13.1	25.6	1.1	1.5	4.0	6.8	644	593
24	28.4	16.9	51.6	45.8	15.1	26.5	1.5	2.0	3.4	8.8	800	739
Total.	35.5	24.8	47.4	43.5	11.9	21.1	1.2	1.9	4.0	8.7	6859	6608

[1] Girls answered with regard to what they believed their sweethearts, brothers, or husbands should do.

[2] Or if invasion threatened.

[3] Includes those who would seek exemption, or perform only non-combatant service.

[4] Within each age the percentages for each sex add to 100.0.

While the female youth show the most pronounced attitudes relative to this question, they also show regularly for each age about twice the proportion of uncertain action as do the male youth.

Wages Received.—The attitude of the employed youth toward their own weekly wages indicated the general extent of dissatisfaction among youth relative to the wages they received. The evidence is clear that as age increased the proportion who believe that they are paid what they are worth declined from fifty-seven per cent at age sixteen to thirty-seven per cent at age twenty-four. Correlatively the proportions who stated they were underpaid increased from thirty-one per cent to forty-nine per cent for the same age range.

The trend by age for a particular occupation or field would throw further light on this question, but in the absence of that the following occupational differences will be cited: Among the female employed youth the groups showing the greatest amount of wage dissatisfaction

were the teachers and the textile factory machine workers, sixty per cent of each stating that they were underpaid. Beauticians and bookkeepers showed the lowest proportions: twenty-seven and twenty-five per cent, respectively. Among the skilled or semi-skilled male employed youth the textile factory machine workers showed the greatest dissatisfaction (seventy-one per cent) while the clothing factory machine workers and truck drivers ranked lowest (thirty-nine per cent).

Economic Value of Schooling.—The effect of experience on youth's attitude toward the economic value of his own schooling is clearly shown by the data. While the proportion who saw little or no economic value in his schooling increased by age from twenty to thirty per cent, those who thought it was of great or considerable value dropped from sixty-three to forty-eight per cent, most of which took place between the ages of sixteen and nineteen. This suggests the disillusioning effect of experience in the employment market.

Sex Education in the Elementary Grades.—The answers to the question "If sex education should be given in the school at what level should it begin?" showed that a fair proportion of both male and female youth in favor of beginning it somewhere in the elementary-school grades. The percentages of youth in favor increased from fifteen to twenty-three for the male group, and from eighteen to thirty-two for the female group, as age increased from sixteen to twenty-four years. At every age the proportion in favor was from three to ten per cent higher among the female youth than among the male youth, revealing a sex as well as an age difference with regard to this question.

INTERPRETATION

Approximately constant proportions appear to be associated with certain verbal opinions and beliefs which can be taken on at a relatively youthful age. This can be done because no direct personal behavior is involved in holding the belief. It may be that many social and economic opinions and beliefs acquire adult proportions of acceptance even before the teen age is reached. Opinions which are widely discussed and deliberately taught in the home and elsewhere are likely to be of this type.

The fact that certain verbal attitudes are held as commonly among sixteen-year-olds as among twenty-four-year-olds suggest the need of determining the trend of the line beyond these limits. The formation of social attitudes of national and international importance is not

limited to the older or to the so-called "mature" ages over twenty-one. They are as prominently held by those in their early teens. If it can be shown that certain basic social and economic attitudes are formed in the four, six, or eight years prior to the age of sixteen, then the challenge to parents and to teachers is indeed great.

The attitudes whose proportions change with age are not the verbal opinions but the more behavioristic patterns which, obviously, involve overt behavior relative to that of others. Possibly this difference reflects the growing change in values which youth acquire as they compare social and economic realities which they meet with childhood teachings and taboos. The attitudes toward drinking and sex education in the elementary grades appear to be reactions against mores of long standing. The decline in church attendance may grow out of increasing social and economic obligations and the corresponding change in interests arising therefrom. To what extent it represents a breaking away from parental control, to what extent it means that youth formerly received social and emotional benefits from the church which they are now obtaining in other ways, we cannot say.

One possible explanation of the change in attitude toward war, slight as it is, may be that the older youth are more conscious of something to live for. It is well known that recruiting appeals draw the younger unemployed and otherwise footloose male much easier than his older married and employed brother. While there appears to be an appreciable increase in the opposition to war as the age of youth increases, still this group is relatively small, accounting for no more than one-quarter of the whole.

Economic security increasingly becomes a problem of major concern to these youth as they grow older. Their attitude toward the economic value of their schooling, and that of employed youth toward their own wages, presumably show the effects of experience in the economic world. In any case, it appears that as age increases the tendency is to put a higher evaluation on oneself and a lower rating on schooling.

HYPOTHESES TO BE TESTED

The foregoing data indicate the necessity of investigating basic attitudes at all the various ages from childhood to adulthood in order to determine when they begin, what course they take, and how stable they remain. We present herewith several working hypotheses which may be used as the basis of further investigation:

(1) The line of proportion for many opinions and beliefs will tend to be constant with age beginning at some age prior to sixteen years.

(2) The line of proportions for many kinds of overt behavior will tend to vary with age beginning at some age prior to sixteen years.

(3) The line of proportions for certain kinds of overt behavior of a highly emotional character will tend to be constant with age beginning at some age prior to sixteen years. This might be true of attitudes toward sex behavior, toward murder, toward the Negro.

(4) The rise and fall of belief or behavior in a given age range will be indicated by a single-peaked line of proportions. Literal belief in Santa Claus may be an example.

(5) A relatively sharp break in the line of proportions such as represented by a regular or by an inverted S shaped curve indicates a significant shift in attitudes as between one age group and another. An example of such an attitude would be drinking, which appears to develop during late adolescence; possibly certain sex attitudes develop in a similar way. Any differences between the attitudes of the older and the younger generations would be shown in this way.

(6) The line of proportions may have at least two maximum points, with a minimum point between these two. An example of this might be church attendance which is characteristic of the very young and the elderly more than of the group between. Economic insecurity may follow such a curve since youth and the aged may be the least secure.

AN ANALYSIS OF QUALITIES ASSOCIATED WITH LEADERSHIP AMONG COLLEGE STUDENTS

E. C. HUNTER

Converse College

AND

A. M. JORDAN

University of North Carolina

INTRODUCTION

The history of civilization is in large part the biography of great men. So universal and enduring is the respect of eminence and greatness among men that personal qualities are often greatly overemphasized. Outstanding achievements and contributions result from altogether too few of those who represent the highest in human development. That the most capable should assume positions of leadership has been an ideal since the time of Plato. Nevertheless, only a meager beginning has been made in developing methods of discovering potential leadership ability and in giving specific training in order that the most able individuals will gravitate to positions of leadership. Early studies of leadership were hardly more than subjective observations. These consisted largely in listing attributes of character and personality which were supposed to be typical of all leaders. Many of the characteristics were so general and complex as to be of little value in the analysis of leadership. Certain investigators approached the problem from the anthropologist's point of view, emphasizing some special ability of the leader as developed when the primitive group faced a crisis. Others have been interested in the social processes of present-day group life, emphasizing the power of the leader in energizing the group. Many studies have attempted to measure qualities purported to be associated with the holding of leadership positions. These studies have emphasized traits of personality, intelligence being the most significant. Some recent attempts have been made in appraising the leadership requirements of definite situations, although it is extremely difficult to maintain a natural setting for the observations. An adequate survey of previous studies of leadership does not fall within the scope of this paper, but perhaps brief mention of a few representative investigations will aid in clarifying both problem and method.

The early and well-known statistical studies of Galton and Woods, supporting the conclusion that genius is hereditary, and those of Ellis, Odin, Cattell, Thorndike, and others, pointing to the environment as a controlling factor in the production of genius need not concern us, except in so far as we may assume a positive correlation between leadership and eminence. Terman,[1] in 1904, made the first noteworthy attempt to place the study of leadership on a scientific basis. He showed that leaders among school children tended to be larger, better dressed, brighter in school work, of more prominent parentage, more daring, more fluent of speech, more attractive, more interested in reading, more suggestible, less emotional, and less selfish than the automatons. Also, intelligence increased rapidly in importance from the second to the eighth grade, and goodness as rapidly fell as a dominant reason given by pupils for choosing certain of their schoolmates for ideals. Caldwell[2] found that high-school pupils listed intelligence, dependability, social adaptability, age, and size in answer to why particular individuals were chosen to lead the class in certain definite situations. In answer to why certain girls were chosen for captains of their gymnasium squads, Nutting[3] obtained such traits as obedience, honesty, "plays fair," "is capable," "able to control her team," dependability, trustworthiness, "knows her business," and "watches what she is doing." The leaders appeared to be above the average in intelligence, popularity, age, physical index, and slightly below in scholarship. Prosh[4] concluded that almost any quality developed to a greater extent in one individual than in his associates was easily recognized by the group. Social aspects of personality, however, appeared to have more influence on the group than other aspects. Gowin[5] showed that leaders in various situations possess physical characteristics which are different from those of lesser leaders and followers in the same situation. Bishops were taller and heavier than

---

[1] Terman, L. M.: "A Preliminary Study in the Psychology and Pedagogy of Leadership." *Pedagogical Seminary*, Vol. XI, 1904, pp. 413–451.

[2] Caldwell, O. W.: "Some Factors in Training for Leadership," National Association of Secondary School Principals. *Fourth Yearbook*, 1920, pp. 2–13.

[3] Nutting, L. R.: "Some Characteristics of Leadership." *School and Society*, Vol. XVIII, 1923, pp. 387–390.

[4] Prosh, F.: "The Basis on Which Students Choose their Leaders." *American Physical Education Review*, Vol. XXXIII, 1928, pp. 265–269.

[5] Gowin, E. B.: *The Executive and His Control of Men.* New York, The Macmillan Company.

clergymen in small towns; university presidents taller and heavier than presidents of smaller colleges; and city superintendents taller and heavier than principals in small towns.

Bowden[1] reported that presidents of student bodies and student councils in forty colleges and universities in the United States possessed a median height of sixty-eight inches, a median weight of one hundred fifty-seven and six-tenths pounds, a low pitched voice on the average, light complexion, dark hair, a median age of twenty-one and five-tenths years, a median IQ of 116, an expansive or ascendant personality, good insight and judgment, and more than average ability to adjust to new situations. Bellingrath[2] found boys elected to positions of leadership in extra-curricular activities to be slightly older and larger than their fellow-students. Girl leaders, however, differed markedly from girls who were not leaders—they were younger, taller, heavier, made better marks in school, came from better homes, and exhibited better school attitudes and habits than the non-leaders. Cowley[3] reported that leaders excelled followers (differences statistically significant) in six traits in three different situations: self-confidence, finality of judgment, motor impulsion, speed of decision in arranging aggressiveness motto cards, speed of decision in filling out a questionnaire, and speed of decision in arranging tact motto cards. Brown[4] found high-school leaders to be selected with reference to intelligence, scholarship, nativity of pupils and parents, economic status, health and physique, diversity of interests, and social status. Also, they tended to choose college preparatory subjects and to enter the professional group of occupations in greater proportions than non-leaders. The vast majority of studies have disclosed that intelligence is a definite part of leadership. In one study it is reported that intelligence has about twice the weight in the choice of leaders that it

---

[1] Bowden, A. O.: "A Study of the Personality of Student Leaders in Colleges in the United States." *Journal of Abnormal and Social Psychology*, Vol. XXI, 1926, pp. 149–160.

[2] Bellingrath, G. C.: *Qualities Associated with Leadership in Extra-curricular Activities of the High School.* Teachers College, Columbia University. Contributions to Education, No. 399, 1930.

[3] Cowley, W. H.: "The Traits of Face-to-Face Leaders." *Journal of Abnormal and Social Psychology*, (October–December, 1931); also "Three Distinctions in the Study of Leaders." *Journal of Abnormal and Social Psychology*, Vol. XXIII, (July–September, 1928), pp. 144–157.

[4] Brown, M.: *Leadership Among High-School Pupils.* Teachers College, Columbia University. Contributions to Education, No. 559, 1933.

has in the choice of friends.[1] J. K. Folsom, commenting on Chapin's study of leadership at Smith College,[2] states that extensive and intensive activity go together. "The more activities one is engaged in, the more offices he is likely to hold, and the more his scholarship is apt to be superior."[3] Summarizing a study in Detroit, Murphy and Murphy[4] showed that even young children value ability, achievement, strength, energy, endurance, liveliness, and imagination in their leaders. In a recent study Partridge[5] reported that outstanding leaders in situations of the Boy Scout nature excelled their fellows in every characteristic measured. Intelligence, dependability, appearance, athletic ability, and voice were associated with leadership in the order named. No evidence was presented to show that there were typical leaders. The leaders differed greatly in personality, physical appearance, and general intelligence.

Though certain qualities and characteristics may be associated with leadership, it is not a simple trait and does not operate consistently. It cannot be defined adequately. It appears to be a function of the situation, depending upon attitudes, disposition, and ability to influence others to act and think in a certain way. According to Goodenough, the most important attribute of the successful leader of any age or level of development is probably the "ability to recognize the special abilities and limitations of others, together with the versatility in devising rôles into which these characteristics will fit."[6] Furthermore, any useful criterion of leadership should distinguish between "leadership" and "headship." The latter may occupy a position of leadership because of prestige.

The present study is limited for the most part to an intensive consideration of qualities associated with leadership in one situation—student life in a southern State University with a student enrollment of about three thousand. The investigation extended over one school

---

[1] Detroit Teachers College, "How Children Choose Friends." *First Annual Report of the Sociey for the Scientific Study of Character*, 1929.

[2] Chapin, F. S.: "Measuring the Volume of Social Stimuli." *Social Forces*, Vol. IV, 1926, p. 479, and "Leadership and Group Activity." *Journal of Applied Sociology*, Vol. VIII, 1924, p. 141.

[3] Folsom, J. K.: *Social Psychology*, pp. 596–597.

[4] Murphy, G. and Murphy, L.: *Experimental Social Psychology*, pp. 358, 402–407.

[5] Partridge, E. D.: *Leadership Among Adolescent Boys*. Teachers College, Columbia University, Contributions to Education, No. 608, 1934.

[6] Goodenough, F. L.: *Development Psychology*, p. 277.

year and consisted of an analysis of a relatively large number of traits and characteristics of college students who were recognized as leaders in various campus activities. A group of typical non-leaders or average students was chosen from the same student body to serve as a control group. The qualities and characteristics measured for each group for purposes of comparison included age, height, weight, general physique, residence, political party preference, church affiliation, fraternity affiliation, parental education, vocational and occupational preference, choice of college curriculum, self-help participation, intelligence, vocabulary, college scholarship, maturity of interests, neurotic tendency, self-sufficiency, introversion-extroversion, dominance-submission, and attitudes toward the Negro, war, economic and labor problems, social life, government issues, and religion.

TECHNIQUES AND PROCEDURES

a. Number and Selection of Groups

Three techniques were employed in the selection of the group of leaders: Student ratings, faculty ratings, and leadership records. Twenty-two leading national fraternities on the campus were invited by letter to coöperate in an investigation, the purpose of which was stated to be an analysis of qualities and abilities of students who were recognized by their fellows as outstanding leaders and who possessed an unusual degree of leadership ability. Each fraternity group was requested to select by secret ballot its two most able leaders, giving consideration so far as possible to the following qualities: Sociability, participation in extra-class activities, scholarship, effectiveness in dealing with people, organizing ability, honors and distinctions, and the number of positions of leadership held in the fraternity and on the campus prior to the time of the study. The fraternities coöperated one hundred per cent, sending in immediately the ballots on which members had indicated their choices of leaders. The votes were tabulated and thus two of the most outstanding men in each fraternity were selected. Approximately fifty students were chosen by this method against which faculty ratings and leadership records were checked.

The various deans of schools and department heads in the University, certain faculty members whose contacts with students were unusually rich, and students then occupying important campus offices were sent blanks on which they were asked to list the names of ten

upperclassmen possessing high degrees of leadership. In order to have various types of leaders represented in the ratings, the same qualities were listed on the blanks as had appeared on the fraternity ballots. The faculty ratings were then checked against the fraternity ratings. In this way partial correction was made for the possible social bias resulting from the selection of the social fraternities. It must be remembered, however, that the fraternity man, by virtue of his election to membership in the organization, probably represents more than average leadership in some respects. For this reason it was deemed justifiable to solicit the aid of the fraternities in choosing the leadership group. The second technique, moreover, yielded additional leadership candidates from non-fraternity men.

The third and last step in the selection of leaders provided objective and reliable evidence against which were checked the more subjective ratings in steps one and two. Provision was made on one of the measuring instruments (A Test of Social Attitudes) for students to list all activities of an extra-class nature in which they had participated in high school or in the first two or three years in college. The date was given as well as all positions of leadership held in connection with the activities. Among the most important activities listed by students were sports, publications, debating, dramatics, student government, class offices, social orders, music, Y. M. C. A., and a variety of clubs. These leadership records listed by the students themselves were carefully checked against the fraternity and faculty ratings. A student was included in the leadership group provided his leadership record supported and was supported by faculty and fraternity or student ratings. At length, a group of eighty-two students, designated the leadership group, was selected. It comprised three sophomores, twenty-eight juniors, forty seniors, and eleven law students. From the leadership group the twenty-five students possessing the highest ratings for leadership ability were chosen and labeled the outstanding leadership group.

Non-leaders, comprising one hundred three students, were selected to represent average students who did not show any marked degree of leadership. This group included four sophomores, fifty-five juniors, forty-three seniors, and one law student. The two groups thus were practically equivalent in class status at the University.

(*b*) *Sources and Description of Data*

The following instruments were used to obtain objective measures for the two groups: (1) The American Council Psychological Examina-

tion yielded the intelligence test data. This test had been administered to the students as freshmen and their records were available in the Dean's office. (2) The North Carolina High-school Senior Examination provided measures of general high-school accomplishment. This test had been administered to the students prior to this investigation and the records of a large percentage of them were available in the Dean's office. (3) College grades were obtained from the office of the Registrar. (4) The Inglis Vocabulary Test of one hundred fifty carefully selected words was used in the study as a result of recent investigation showing the value of the test in contributing to the differentiation of various degrees of executive success.[1] (5) The Strong Vocational Interest Blank yielded ratings for all students in nine occupations, including architect, C. P. A., chemist, engineer, journalist, lawyer, personnel manager, physician, and city superintendent. Also, the blank provided a measure of maturity of interests. (6) The Bernreuter Personality Inventory yielded measures of neurotic tendency, self-sufficiency, introversion-extroversion, and dominance-submission. (7) The Test of Social Attitudes,[2] constructed especially for this study, yielded attitude scores on social, racial, economic, governmental, political, and religious issues. It was designed to measure various degrees of certainty, indifference, or doubt existing among students with respect to ninety-four questions. The statements were so framed that a measure of liberalism-conservatism on each item and on each section of the test could be obtained. Also, a total liberalism-conservatism score was computed. The following are sample items taken from the test:

2 1 0 −1 −2 Negroes should never be placed in authority over white people. (13)

2 1 0 −1 −2 The majority of labor strikes in this country are justified. (44)

DIRECTIONS

Examine each statement carefully and draw a circle around the number in the margin which best indicates your judgment, as follows:

(2) 1 0 −1 −2 Encircle "2" if you are strongly convinced that the statement is true.

2 (1) 0 −1 −2 Encircle "1" if you are fairly certain the statement is true, that it is true in large degree, that it is more true than false.

2 1 (0) −1 −2 Encircle "0" if you are undecided, uncertain, and have no information or conviction on the statement.

---

[1] O'Connor, Johnson: "Vocabulary and Success." *Atlantic Monthly*, February, 1934.

[2] To be published later.

2 1 0 (-1) −2 Encircle "−1" if you have a feeling against the statement and think that it is more false than true, that it is false in large degree.

2 1 0 −1 (-2) Encircle "−2" if you are strongly convinced that the statement is false, that no reasonable person would consider it true.

(The number which is encircled indicates the degree of liberalism-conservatism for that item.)

RESULTS

Table I presents comparisons of two leadership groups with non-leaders with regard to sixteen traits and capacities. The number of cases, the mean, the probable error of the mean, and the standard deviation are given for each of the three groups in each trait. Comparisons of each of the leadership groups with the unselected or non-leadership group are indicated by the mean difference and the probable error of the difference between the means. The critical ratio (CR), the ratio of the mean difference to the probable error of the difference, is shown for each comparison. This measure affords a test of the significance of the difference between each of the leader groups and the non-leaders. A difference is not considered statistically significant unless it is four times the size of its probable error; however, McGaughy and others have shown that a difference between means is significant when it is "at least three times as great as the probable error of that difference."

SUMMARY OF FINDINGS

(1) *Age.*—Leaders were significantly younger than non-leaders of the same class status as indicated by a mean difference of 5.9 months and a critical ratio of 3.04. A group of twenty-five outstanding leaders averaged nine and eight-tenths months younger than non-leaders of the same class status, the critical ratio being 4.16, as shown in Table I.

(2) *Physique.*—The difference in weight between leaders and non-leaders was significant, the non-leaders averaging seven and nine-hundredths pounds heavier. With respect to height and general health conditions no significant differences were found between leaders and non-leaders. A slight tendency for leaders to be taller and to possess fewer physical defects was noted.

(3) *Residence.*—Approximately the same proportion of leaders as non-leaders were out-of-State students. The percentages were

TABLE I.—COMPARISON OF LEADERS (*L*) AND OUTSTANDING LEADERS (*OL*) WITH NON-LEADERS (*NL*) IN REGARD TO VARIOUS TRAITS

Trait	Group	*N*	Mean	PE_m	SD	Mean diff.	$PE_{diff.}$	CR
Age (months)	*L*	82	255.8	±1.13	15.2	−5.9	±1.94	3.04
	OL	25	251.9	±1.76	13.1	−9.8	±2.36	4.16
	NL	103	261.7	±1.58	25.5			
Weight (lbs.)	*L*	82	157.56	±1.24	16.6	−7.09	±1.87	3.79
	OL	25	156.50	±2.47	18.3	−8.15	±2.84	2.87
	NL	103	164.65	±1.40	21.1			
Intelligence (American Council Psychological test percentile).	*L*	68	76.40	±1.63	20.0	20.5	±2.50	8.20
	OL	24	84.60	±1.95	14.2	28.7	±2.72	10.55
	NL	77	55.90	±1.90	24.6			
Vocabulary (Inglis test)	*L*	82	123.7	±0.82	11.1	21.9	±1.51	14.50
	OL	25	129.1	±1.69	12.6	27.3	±2.11	12.90
	NL	103	101.8	±1.27	19.2			
Scholarship (grades)	*L*	82	2.53	±0.07	0.91	0.39	±0.09	4.43
	OL	25	2.29	±0.12	0.86	0.63	±0.13	5.04
	NL	103	2.92	±0.05	0.79			
Interest maturity (Strong blank percentile).	*L*	82	71.8	±1.95	26.2	16.0	±2.71	5.92
	OL	25	81.3	±2.74	20.3	25.5	±3.32	7.68
	NL	102	55.8	±1.88	28.2			
Neurotic tendency (Bernreuter percentile).	*L*	82	43.9	±2.12	28.5	−7.7	±2.73	2.82
	OL	25	44.3	±3.88	28.8	−7.3	±4.24	1.72
	NL	103	51.6	±1.73	26.1			
Self-sufficiency (Bernreuter percentile).	*L*	82	55.0	±1.92	25.8	14.0	±2.56	5.47
	OL	25	62.7	±3.55	26.3	21.7	±3.93	5.52
	NL	103	41.0	±1.68	25.3			
Introversion-extroversion (Bernreuter percentile).	*L*	82	43.2	±2.16	29.0	−3.6	±2.80	1.28
	OL	25	44.9	±3.76	27.9	−1.9	±4.16	0.46
	NL	103	46.8	±1.78	26.8			
Dominance-submission (Bernreuter percentile).	*L*	82	66.5	±2.12	28.5	11.4	±2.80	4.07
	OL	25	77.8	±3.26	24.2	22.7	±3.74	6.07
	NL	103	55.1	±1.84	27.8			
Attitude toward Negro (Hunter test-raw score).	*L*	82	8.65	±1.13	15.2	9.35	±1.50	6.25
	OL	25	17.1	±1.45	10.7	17.8	±1.76	10.11
	NL	103	−0.7	±0.99	14.9			
Attitude toward war (Hunter test-raw score).	*L*	82	6.0	±0.76	10.2	1.75	±0.90	1.94
	OL	25	8.9	±1.47	10.9	4.65	±1.55	3.00
	NL	103	4.25	±0.50	7.6			
Attitude toward labor-economics (Hunter test-raw score).	*L*	82	8.2	±0.70	9.4	1.3	±0.92	1.41
	OL	25	12.2	±1.44	10.7	5.3	±1.56	3.40
	NL	103	6.9	±0.59	8.9			
Attitude toward social life and convention (Hunter test-raw score).	*L*	82	6.7	±0.45	5.9	1.1	±0.58	1.90
	OL	25	7.3	±0.86	6.4	1.7	±0.93	1.83
	NL	103	5.6	±0.36	5.4			
Attitude toward government issues (Hunter test-raw score).	*L*	82	1.9	±0.48	6.4	1.2	±0.61	1.98
	OL	25	4.6	±1.04	7.6	3.9	±1.10	3.50
	NL	103	0.7	±0.38	5.8			
Attitude toward religion (Hunter test-raw score).	*L*	82	−2.1	±0.50	6.6	−0.1	±0.62	0.02
	OL	25	0.4	±1.10	8.1	2.4	±1.16	2.07
	NL	103	−2.0	±0.37	5.6			
General attitudes (Hunter test-total raw score).	*L*	82	31.2	±3.10	41.6	14.2	±3.80	3.74
	OL	25	51.8	±5.95	44.1	34.8	±6.34	5.33
	NL	103	17.0	±2.19	32.9			

twenty-four and three-tenths and twenty-three and two-tenths, respectively. Of the twenty-five outstanding leaders only twelve per cent resided outside of the State.

(4) *Political Party Preference.*—Little or no relationship was found between leadership and political party preference. There was, however, a tendency for leaders to favor no political party.

(5) *Church Affiliation.*—A significantly larger proportion of Episcopalians than that of any other denomination was found among the leaders. The percentage of Episcopalians among the leadership group was twenty-eight and one-tenth while that for the University at large was only fifteen and nine-tenths. Also, sixteen per cent of the outstanding leaders and only seven and eight-tenths per cent of the non-leaders indicated no church affiliation.

(6) *Fraternity Affiliation.*—Eighty per cent of the leaders and forty-eight per cent of the non-leaders were fraternity members. Apparently, it is fair to conclude from this study that the social fraternity is an agency of selection of leaders on the college campus.

(7) *Education of Parents.*—The proportion of parents of leaders who had attended college was more than twice that of parents of non-leaders. Mothers of leaders had attended college in greater proportions than had fathers. For the three groups, leaders, outstanding leaders, and non-leaders, the percentages of fathers who had attended college were forty-six and four-tenths, forty-eight, and twenty-three and three-tenths, respectively. Corresponding percentages of mothers who had attended college were fifty-seven and three-tenths, seventy-two, and twenty-seven and two-tenths.

(8) *Occupational Status of Parents of Leaders and Non-leaders.*—The proportion of leaders' parents belonging to the professional classes was definitely larger than that of non-leaders' parents. The leadership group occupied a higher position in the scale of parental occupational status than did the non-leaders.

(9) *Vocational Preferences of Leaders and Non-leaders.*—Leaders showed a great deal more restriction in their vocational choices than did non-leaders. Approximately seventy-five per cent of the leaders had chosen law or journalism for their life's work. On the Strong Interest Blank, leaders made substantially higher interest ratings in law and journalism than did non-leaders.

(10) *Self-Help Participation.*—Leaders worked on the average a smaller number of hours per week in self-help activities than non-leaders. The study revealed that thirty-nine and eight-tenths per cent

of the non-leaders, thirty-nine per cent of the leaders, and forty-eight per cent of the outstanding leaders averaged weekly twenty-two and two-tenths hours, sixteen and one-tenth hours, and thirteen and seven-tenths hours, respectively, for four years.

(11) *Intelligence.*—Table I reveals that the intelligence of the leaders was found to be significantly higher than that of non-leaders, the mean percentile difference being 20.5 points with a probable error of the difference of 2.5. Outstanding leaders ranked 28.7 percentile points above the non-leaders.

(12) *Vocabulary.*—Significantly higher scores were made by leaders than by non-leaders as measured on the Inglis Vocabulary Test. Outstanding leaders showed even greater superiority in vocabulary as shown in Table I.

(13) *Scholarship.*—Leaders made significantly higher grades on courses completed the year immediately previous to the investigation than did non-leaders. Outstanding leaders made even higher grades than the lesser leaders. Table I shows the scholarship index for leaders to be 2.53, for outstanding leaders to be 2.29, and for non-leaders to be 2.92. ($A = 1$, $B = 2$, $C = 3$, etc.) Also, on the North Carolina High-School Senior examination, administered in the senior year of high school, the leaders made substantially higher scores than did the non-leaders.

(14) *Interest Maturity.*—As measured by the Strong Vocational Interest Blank, the leaders were found to be significantly more mature in vocational interests than non-leaders. As shown in Table I leaders were sixteen per cent and outstanding leaders twenty-five per cent more mature in interests than non-leaders even though the latter were actually from five to ten months older than leaders.

(15) *Neurotic Tendency.*—As measured by the Bernreuter Personality Inventory no real difference was found between leaders and non-leaders, or between outstanding leaders and non-leaders with regard to neurotic tendency. There was, however, a slight tendency for leaders to be less neurotic than non-leaders.

(16) *Self-Sufficiency.*—Leaders were found to be more self-sufficient than average students of the same educational status. Outstanding leaders showed significantly higher percentile ratings in self-sufficiency than did the non-leaders.

(17) *Introversion-Extroversion.*—No significant difference was found between leaders and non-leaders with regard to introversion tendency. On the whole, the students in this study appeared to be slightly less

introverted than average individuals of the same educational status. The percentile ratings of the two leadership groups on this scale were almost identical with those on the scale for neurotic tendency, showing that similar scores on the two scales bear the same significance.

(18) *Dominance-Submission.*—Leaders were found to be significantly more dominant in face-to-face situations than non-leaders. The percentile ratings of the leaders and outstanding leaders were eleven and four-tenths and twenty-two and seven-tenths higher than that of the non-leaders. (See Table I.)

(19) *Attitudes toward the Negro.*—As determined by the Test of Social Attitudes, the leaders were significantly more liberal toward the Negro than non-leaders. (The term *liberal* as used in this study means an attitude or disposition favorable to an improved social, political, and economic status for the Negro.) Outstanding leaders were substantially more liberal than lesser leaders as shown in Table I.

(20) *Attitudes toward War.*—A slight tendency for the leaders to score more liberally (pacifistic attitude) toward war issues than non-leaders was discovered by the attitudes test. Outstanding leaders showed even higher liberal scores, but the difference was hardly statistically significant.

(21) *Attitudes toward Capital and Labor.*—Leaders, especially outstanding leaders, were found to surpass non-leaders in showing more sympathy for labor than for capital. The difference between the scores of leaders and non-leaders was not significant, however.

(22) *Attitudes toward Social Life and Convention.*—No significant difference was revealed by the attitudes test between leaders and non-leaders with respect to their attitudes toward existing social conventions. On the whole, the students exhibited a liberal tendency, or an attitude favorable to experimentation for the purpose of advancing human happiness and welfare.

(23) *Attitudes toward Government.*—No real difference was found between leaders and non-leaders with regard to their attitudes toward government issues. Outstanding leaders made higher liberal scores than did non-leaders, but the difference was not significant. (A liberal score was taken to signify an attitude favorable to changes in government for the common welfare and one unfavorable to a capitalistic government which supports vested interests.)

(24) *Attitudes toward Religion.*—No significant difference was found between leaders and non-leaders with regard to attitudes toward religious issues. There was a tendency for all groups to score con-

servatively, that is, to imply support of present church and religious practices.

(25) *General Attitudes.*—On the attitudes test as a whole, the leaders made significantly more liberal scores than did the non-leaders. The difference was even more pronounced between outstanding leaders and non-leaders as shown in Table I. Leaders, outstanding leaders, and non-leaders scored 31.2, 51.8, and 17.0, respectively. Intercorrelations revealed that leaders were more consistent than non-leaders in their attitudes toward various aspects of life. Leaders' attitudes toward the Negro were found to be more predictive of their attitudes toward war, economics and labor, and government than was true of non-leaders. In religion, however, non-leaders were found to be more consistent. A knowledge of their religious attitudes was somewhat more predictive of their attitudes toward the Negro, toward social life, and toward government issues than was true of the leaders. Thus in the realm of religion, it appears that leaders showed more doubt and uncertainty than did the non-leaders.

THE TYPICAL LEADER IN THIS STUDY

The typical leader in this study was younger, lighter, and possessed fewer physical defects than his classmates. He was far superior in general intelligence, scholarship, and vocabulary. His parents were college trained. Law and journalism were his chief vocational interests. He was a member of a social fraternity and possessed more maturity in interests, more self-sufficiency, and more dominance in face-to-face situations than his fellow students. He showed more "liberal" attitudes than his fellows toward the Negro, toward war, toward labor and economic issues, toward social convention, and toward governmental questions. He was more disposed than his fellows to regard as desirable definite changes in social institutions purporting to advance social welfare. He was more strongly opposed to the *status quo* and to the doctrine of class or vested interests than were his classmates. He was sympathetic with the interests of labor. His attitude was more strongly pacifistic than that of his fellows. He was more accurately informed, read more carefully, offered more resistance to suggestion, used more caution in response to extreme statements, and exercised more discriminating judgment than was true of his classmates and fellow students.

COLLEGE GRADES MEAN SOMETHING

J. B. SHOUSE

Marshall College, Huntington, W. Va.

1

The evidence that a number of people, scoring the merit of some particular piece of work—an examination paper, for example—will (unless quite objective criteria are used) so disagree as to stamp their individual judgments as ill-founded, is accepted as conclusive. Given judges enough, it appears that the paper may be assigned almost every possible rating. In such case there is just one variable, the standards of the judges, and all disagreements as to value are manifestations of that variable.

In comparing grades made in different courses by a student working under different instructors, we have two variables. One variable is the element of faculty standards of judgment; the other is the element of student reaction to different courses and different teachers. There is no reason to believe that the second variable is less influential than the first.

With two variables in operation in any situation it seems logical to expect to find a more pronounced scattering of grades than is found in the rating of a single piece of work by several judges, the case of a single variable. On the other hand, when a student pursues a course for an entire semester, the instructor has opportunity to score a more extensive showing of quality than is found in one examination. I have sought to determine whether, in that case, there is greater, or less, predictability of outcome than there is when a single paper is graded. As a working hypothesis I suggest that grading a semester's work is quite another thing than evaluating a single performance, and that to final grades for the semester there attaches a degree of significance not always conceded.

In pursuit of evidence bearing upon the hypothesis I examined the records of fifty college seniors during the school year 1936–1937. Each of these students had just completed his seventh consecutive semester in this institution; some had, in addition, done some Summer school work. Names were taken in alphabetical order; there is no reason to suspect any selection in the listing of the names. All final grades for each student were tabulated to show number of courses taken, number of hours of work represented by those courses, number

TABLE I.—PERCENTAGE DISTRIBUTION OF GRADES OF FIFTY COLLEGE SENIORS

Student	Per cent courses graded					Per cent hours graded					Per cent courses in consecutive grades	
	A	*B*	*C*	*D*	*F*	*A*	*B*	*C*	*D*	*F*	Two grades	Three grades
1	27	46	25	0	2	26	50	21	0	3	73	98
2	15	36	47	2	0	14	39	46	1	0	83	98
3	83	15	2	0	0	83	14	3	0	0	98	100
4	36	34	24	3	3	35	36	25	1	3	70	94
5	0	20	50	23	7	0	18	52	22	8	73	93
6	2	29	62	7	0	4	29	60	7	0	91	98
7	56	27	17	0	0	57	27	16	0	0	83	100
8	44	36	20	0	0	43	37	20	0	0	80	100
9	11	57	32	0	0	10	57	33	0	0	89	100
10	20	43	30	7	0	18	44	29	9	0	73	93
11	29	51	20	0	0	32	47	21	0	0	80	100
12	46	26	24	4	0	50	27	19	4	0	72	96
13	60	38	2	0	0	60	37	3	0	0	98	100
14	0	25	45	27	3	0	26	45	27	2	72	97
15	51	39	10	0	0	52	37	11	0	0	90	100
16	3	18	56	20	3	3	20	53	22	2	76	94
17	16	26	56	2	0	8	27	63	2	0	82	98
18	17	49	34	0	0	19	46	35	0	0	83	100
19	0	32	52	16	0	0	32	52	16	0	84	100
20	9	62	29	0	0	12	60	28	0	0	91	100
21	12	40	36	12	0	13	42	34	11	0	76	88
22	11	35	46	8	0	12	34	46	8	0	81	92
23	25	46	29	0	0	29	38	33	0	0	75	100
24	65	22	13	0	0	63	23	14	0	0	86	100
25	6	59	33	2	0	6	59	34	1	0	92	98
26	4	21	63	10	2	5	22	62	9	2	84	94
27	18	47	33	2	0	21	50	28	1	0	80	98
28	0	14	77	7	2	0	12	76	9	3	91	98
29	7	50	43	0	0	9	50	41	0	0	93	100
30	11	55	34	0	0	11	56	33	0	0	89	100
31	82	18	0	0	0	86	14	0	0	0	100	
32	62	31	7	0	0	65	31	4	0	0	93	100
33	17	35	42	6	0	17	32	44	7	0	77	94
34	89	11	0	0	0	92	8	0	0	0	100	
35	2	33	51	14	0	1	31	52	16	0	84	98
36	7	38	45	10	0	7	41	44	8	0	83	93
37	2	21	63	12	2	3	20	63	12	2	84	96
38	39	48	11	2	0	38	49	11	2	0	87	98
39	0	17	57	26	0	0	16	55	29	0	83	100
40	27	62	6	5	0	29	60	6	5	0	89	95
41	2	9	63	20	6	1	3	65	23	8	83	92
42	0	16	48	22	14	0	13	47	23	17	70	86
43	9	30	33	22	6	5	22	39	29	5	63	85
44	4	5	68	23	0	4	5	67	24	0	91	96
45	11	34	47	8	0	12	35	47	6	0	81	92
46	2	12	63	14	9	3	8	63	16	10	77	89
47	2	7	53	33	5	2	3	54	37	4	86	93
48	7	19	67	7	0	6	13	72	9	0	84	91
49	7	12	44	35	2	7	12	46	34	1	79	91
50	51	42	7	0	0	54	42	4	0	0	93	100

of courses with grade *A*, *B*, etc., number of hours with grade *A*, *B*, etc.; to save space these data are not presented here. From the original data another table (presented here as Table I) was derived; it shows percentage distribution of grades.

The reading of Table I may be illustrated by following through the case of Student 1. This student received the grade of *A* in twenty-seven per cent of his courses, *B* in forty-six per cent, *C* in twenty-five per cent, *D* in no course, *F* (failure) in two per cent of his courses. He was marked *A* in twenty-six per cent of the hours of work he had taken, *B* in fifty per cent, *C* in twenty-one per cent, *D* in none, *F* in three per cent of his hours of work. In two consecutive grades (*A* and *B*, in this instance) he received seventy-three per cent of all his grades; in three consecutive grades (*A*, *B*, *C*) he received ninety-eight per cent of all his grades.

Table I supports the following statements: (1) At the end of the seventh college semester, one-third of all students (seventeen of fifty) had at least sixty per cent of their grades in one letter; two-thirds of the same students (thirty-three of fifty) had at least fifty per cent of all grades in one letter; (2) it is the very rare student (only one of these fifty) who does not get at least seventy per cent of all grades within the range of two letters; more than two-thirds (thirty-six of fifty) had at least eighty per cent of all grades within the range of two letters; (3) it is the exceptional student (only four of these fifty) who gets less than ninety per cent of all grades in three consecutive letters; we found about two out of five students (nineteen of fifty) who had placed all grades within the range of three letters.

One may reasonably conclude that there is a very considerable degree of consistency in final grades of college students. Each grade depends on two variable factors; could we eliminate one of these (the variable performance of the student in different courses) it is obvious that the consistency of grades would be greatly increased. Consequently, it seems fair to college teachers generally to say that probably any one of them gives to a student's whole performance for a semester approximately the same grade that any other one of them would give to the same quality of performance.

It follows that one is justified in taking rather seriously any student's final grade in any course as indicative of what the student actually did under the existing conditions of work, whether that performance may or may not be regarded as indicating the student's probable performance under quite different conditions; life performances in later years, for example.

2

During the first semester, 1937–1938, I had under my instruction in a certain course two groups of students, forty-eight of whom completed the course. Final grades were distributed as follows: *A*—2; *B*—7; *C*—28; *D*—9; *F*—2. These grades were derived from point scores; five hundred sixty points could be earned through the performance of varied and frequent exercises; seventy per cent of the maximum number of points were required for *D*, seventy-seven and one-half per cent for *C*, eighty-five per cent for *B*, ninety-two and one-half per cent for *A*. On four occasions during the semester points were summed and grades determined; roughly speaking, these occasions came at the several month ends of the semester. The maximums of points available at the several times were, respectively, one hundred, two hundred fourteen, four hundred seven, five hundred sixty. The record was cumulative. Some of the exercises were objective in character, but the scoring of most of them depended on subjective evaluations in some degree.

Each student, then, had three chances to alter his grade after it was first calculated. That is, for forty-eight students it was possible to have one hundred forty-four changes of grade. Actually only forty-four changes of grade occurred, with thirty-three students involved, the other fifteen students getting the same grade on each occasion. On second grading twenty-five students had different grades than those received at first; the other twenty-three retained first grades. On third rating thirteen students had different grades than those received on second grading; the other thirty-five retained second grades. On fourth grading only six students had different grades than on third grading; the other forty-two retained third grades.

With progressive reduction of changes from twenty-five to thirteen to six, it appears that the final grades of the forty-eight students approximate stabilized positions; had a fifth scoring been possible a month later than the fourth, we may assume there would have been no changes. The final grades, then, are relatively true indices of the quality of work so far as it could be determined by the instructor. That is, through frequent performances of varied kinds the students had opportunity to overcome the statistical effect of variability of performance. Simultaneously, the instructor, by utilizing several varieties of measuring techniques, by fairly frequent measurings, and by means of cumulative records, overcame the statistical effect of variability of his standards of judgment.

There is here no evidence to show that any persistent error of the instructor was overcome, such error as tendency to undue severity or liberality, for example. We have here a simple application of the principle that, while the average of only a few items is very sensitive to the inclusion of additional items, an average based on many items is not noticeably dislocated by the inclusion of more items. What is contended is that that instructor is wise who does not rest the student's grade on slender basis. It takes a good many ratings, apparently, to yield a grade that may be regarded as stabilized, so far as the work in a single course is concerned. The evidence says that the first few ratings did not yield an average that could be regarded as truly representing the quality of the students' work. It is further contended that the instructor finds a reliable basis of grading when he uses many scorings.

Let us assume that any other teacher, making independent ratings on this same work, and using his own series of devices, would have found his grades shifting gradually throughout the semester as did this instructor's. That is, it is reasonable to believe that a comparison of ratings, made by a number of instructors on the basis of a few exercises only, would show all the signs of undependable grading that the scoring of a single set of examination papers shows. The subjective factor would rear its head clearly. But it is also reasonable to believe that, had a number of instructors, each using his own rating devices, followed these students closely throughout the semester, each finally attaining to what appears to be a stabilized grade, a table derived from these final grades would show a very considerable degree of agreement as to the quality of the students' work.

If we can demonstrate anything of the kind, then it follows that a sufficient number of exercises of the essay type, employed now and again, rated at different times, scattered well over the semester, may yield grades as valuable as those obtained on objective examinations. That is, *a battery of essay examinations may prove to be an objective basis of grading college work. At least no one seems to have demonstrated that it is not.* What has been abundantly demonstrated is that a single essay examination does not constitute an objective measurement.

3

During the Summer of 1937 a class of forty students undertook an exercise in reading test papers written by the class. The forty students were seated in rows of ten chairs each. Papers written in

first row were scored by students in third row, for example. Each paper written was read by ten students. Each student, then, read a series of four test papers for each of ten students of the second row from him. For test purposes the text was broken into four parts; each test consisted of eight questions; each answer was to be scored on the basis of a possible five points; each test carried a possible forty points for each student; each student could earn one hundred sixty points on the four tests.

In large majority the members of the class were experienced teachers. The textbook material had been discussed in class before being used as test material. Test papers were read at any hour, according to convenience of reader, but always within a couple of days after being written. Readers were asked to do their scoring independently, but there may have been some comparing of scores in the course of the reading. To a considerable degree, at least, anonymity of writers was preserved. Test questions might fairly be called brief essay exercises. Test periods were fifty minutes, so that approximately six minutes were allotted to each answer; suggestion was made that such a schedule be kept in mind as far as might be convenient in writing.

The data obtained were assembled into a series of tables; illustrative tables only are shown herewith. For Table II I have chosen to display

TABLE II.—SCORES GIVEN BY STUDENTS OF ONE ROW TO PAPERS BY STUDENTS OF ANOTHER ROW, FOURTH TEST
(Writers and Readers Identifiable Only by Key)

Writer	Readers										Totals	Rank
	Q	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*		
A	34	27	35	36	36	29	27	38	36	34	332	6
B	27	30	22	26	36	28	28	37	37	34	305	7
C	29	32	27½	27½	25	27	28	37	37	34	304	8
D	36	34	37	33	37	35	35	38	38	38	361	3
E	38	33	40	33	40	39	34	36	36	36	365	2
F	16	6	12	14	15	5	11	30	31	21	161	10
G	36	34	38	35	39½	38	36	39	39	38	372½	1
H	31	21	27	34	34	29	28	30	32	33	299	9
I	38	35	33	38	33	34	31	38	38	37	355	4
J	37	35	36	35	37	38	35	35	35	28	351	5
Totals.....	322	287	307½	311½	332½	302	293	358	359	333	3205½	

results of reading fourth set of papers written by one group of ten students and read by another group of ten students. Had we used results on first set of papers, the situation would have appeared even more acute.

TABLE III.—RANKS ON FOURTH TEST. DERIVED FROM TABLE II

Writers	Readers									
	Q	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*
A	6	8	5	2	5–6	6–7	9	2–3–4	6–7	5–6–7
B	9	7	9	9	5–6	8	6–7–8	5–6	4–5	5–6–7
C	8	6	7	8	9	9	6–7–8	5–6	4–5	5–6–7
D	4–5	3–4	3	6–7	3–4	4	2–3	2–3–4	2–3	1–2
E	1–2	5	1	6–7	1	1	4	7	6–7	4
F	10	10	10	10	10	10	10	9–10	10	10
G	4–5	3–4	2	3–4	2	2–3	1	1	1	1–2
H	7	9	8	5	7	6–7	6–7–8	9–10	9	8
I	1–2	1–2	6	1	8	5	5	2–3–4	2–3	3
J	3	1–2	4	3–4	3–4	2–3	2–3	8	8	9

Tables II and III show all the phenomena of the typical results obtained by subjective, or partially subjective, grading of a single set of papers by a number of readers. Some are more liberal than are others as shown by total points given; *R*, for example, allowed a total of two hundred eighty-seven points out of a possible four hundred, where *Y* allowed three hundred fifty-nine points on the same ten papers. *Y* confined all scores to the range thirty-one to thirty-nine; *S* let them range from twelve to forty. *A*'s paper, for example, was regarded by *T* as next to best, but by *W* as next to worst in the set. *J*'s paper was regarded as one of best two, good as best, by *R* and as next to worst by *Z*.

Now let us see what the situation is when results of four tests are combined into a single score, so many points on a base of one hundred sixty possible points.

In rather remarkable degree the vast dissimilarity of judgments manifested in Tables II and III has disappeared in Tables IV and V. *A*'s papers are given a ranking of eighth or ninth by every reader, instead of rankings varying from second to ninth. *J*'s papers now are ranked somewhere from fourth to eighth by the several readers; his fourth paper alone was ranked anywhere from first to ninth. So much

TABLE IV.—SCORES GIVEN BY STUDENTS OF ONE ROW TO PAPERS BY STUDENTS OF ANOTHER ROW, FOUR TESTS
(Writers and Readers Same as in Tables II and III)

Writers	Readers										Totals	Rank
	Q	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*		
A	116	99	92½	111½	103½	100	99	126	124	113½	1085	9
B	134½	122	115½	118½	128	121	119	136	138	125	1247½	7
C[1]	131	107	108	105	100½	113	115	140	140	132	1191½	8
D	145	136	141	144	152	145	141	141	147	144½	1436½	2
E	150	127	152	142	152	141	143	144	145	138½	1434½	3
F	60	48	41	50½	34	32	33	99½	102	76½	577½	10
G	151	138	150	146	152½	148½	150	148	151	149	1484	1
H[1]	136	120	127	142	137	119	123	135	141	127	1307	6
I	145	136	133½	147½	134½	133	136	143	146	144	1398½	4
J	144	124	133	136	127	140	140	139	140	124	1347	5

[1] Absent from one test; totals on three tests raised by one-third.

TABLE V.—RANKS ON FOUR TESTS. DERIVED FROM TABLE IV

Writers	Readers									
	Q	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*
A	9	9	9	8	8	9	9	9	9	9
B	7	6	7	7	6	6	7	7	8	7
C[1]	8	8	8	9	9	8	8	5	6–7	5
D	3–4	2–3	3	3	2–3	2	3	4	2	2
E	2	4	1	4–5	2–3	3	2	2	4	4
F	10	10	10	10	10	10	10	10	10	10
G	1	1	2	2	1	1	1	1	1	1
H[1]	6	7	6	4–5	4	7	6	8	5	6
I	3–4	2–3	4	1	5	5	5	3	3	3
J	5	5	5	6	7	4	4	6	6–7	8

[1] Absent from one test; totals on three tests raised by one-third.

has disagreement been reduced by establishing ranks on four papers instead of one paper only that one may readily believe that as many as six tests in the semester would have lead to almost complete agreement as to the ranking of the several students. In such case we would be justified in saying that a battery of essay tests yields results comparable to results of objective tests.

It was pointed out above that student performances themselves constitute a variable. That remark was made respecting work done in different courses. To a less degree would we expect to find evidence of it within the confines of a single course. That the same quality of work will not be manifested by some individuals in all tests in a course is demonstrated clearly by the cases of *A* and *H*. The ten readers of *A*'s fourth paper (see Table II) gave it total points to rank it as sixth in value; *H*'s fourth paper ranked ninth. Table IV shows that the combined scores on four papers from each writer ranked *A* ninth and *H* sixth, however. Evidently *A* did much better, relatively speaking, on his fourth paper than on the other three, while *H* did relatively less well on fourth paper than on the other three.

SUMMARY

I have cited evidence to support the general conclusion that college teachers give more dependable grades than is sometimes admitted. First, I have shown that any student's grades tend to group themselves rather decidedly in spite of developed differences in ability and interest in relation to different courses, and in spite of the factor of different faculty members as graders. Second, I have shown that errors due to subjectivity in the scoring of tests tend to disappear as the instructor combines results of a series of tests. I have approached the latter phase of the problem by two routes: (*a*) I have shown that progressively throughout a semester the student's grade tends to become less subject to change and, therefore, more representative of the average quality of his work, attaining to rather stabilized position on the scale of possible grades if sufficient evidence is incorporated into the grading; (*b*) evidence has also been submitted to indicate that this stabilized position is attained not only as regards one instructor as grader, but that it is also verifiable by other graders scoring the same work. These several bits of evidence mitigate the severity of the judgment often rendered when several scorers rank a single set of examination papers.

THE PROBLEM OF PSEUDO-FEEBLEMINDEDNESS

S. W. BIJOU

Mental Hygiene Clinic and State Hospital, Farnhurst, Del.

In clinical psychology, level of intelligence may be ascertained by either a linguistic or a performance psychometric examination. Language tests require the examinee to read, write, or say something; performance tests require him to do something with his hands. Most clinical psychologists diagnose intelligence by one of the established individual verbal tests, such as the Stanford-Binet, and resort to a well standardized performance scale only when the verbal tests are obviously inadequate. Thus the intelligence of the pre-school child, the illiterate, the foreign born, and those disabled in speech and hearing, is usually determined by a manual test.

Clinical observations and recent investigations indicate that there is still another group of cases whose intellectual capacity is not adequately diagnosed by verbal intelligence tests. These are school children who are (*a*) retarded in grade placement, and (*b*) retarded in reading achievement; and adults who have (*a*) limited educational opportunities, (*b*) non-verbalist propensities, and (*c*) a total reading disability. In other words, this group consists of those who have failed, for one reason or another, to develop their language functions to the level of their native endowment. Jastak[6] found, for example, that "children with special reading disabilities tend to score higher on performance tests than on the vocabulary test if they are mentally stable." The vocabulary test is, of course, the backbone of every verbal test of intelligence. Uhler,[7] in surveying the relationship between intellectual types and forms of maladjustments, discloses that among school children referred to the Delaware Mental Hygiene Clinic who attained a higher performance than verbal IQ, "thirty out of the fifty cases were diagnosed Special Reading Disability on the basis of standard diagnostic tests. Eleven cases were seriously retarded in language functioning, as determined by the Stanford Achievement Tests and Wallin Attainment Scale." Referring to the same group he observes that "if the Binet test had been used alone . . . as a routine procedure, forty-eight per cent of average intelligence children would have been passed as feebleminded." Glanville[3] in interpreting the psychometric pattern of delinquent boys (ages ranged from thirteen to twenty years) concluded "that a large percentage

of the group studied are retarded in their language development far below the level of their native intelligence." When the intelligence of the group is diagnosed according to the Terman Vocabulary or Stanford-Binet, the average boy possesses borderline intelligence; when determined by the performance scale the typical boy is of average intelligence but suffers a marked retardation in language development. Garrett,[2] reviewing a number of studies concerned with the correlation between verbal intelligence tests and school attainment, notes a marked relationship between the two measures which he attributes to the fact that "both tap the same abilities." He argues that the Binet type of tests are concerned with the measurement of verbal capacity, "especially in the ages above seven years. Tests of vocabulary and word meaning, as well as tests of the comprehension of language, and language relationships, play an important part both in school work and in determining the child's IQ."

Because language development influences IQ's derived from verbal intelligence tests, all cases examined psychometrically at the Delaware Mental Hygiene Clinic receive at least one verbal and one performance test of general intelligence. In interpreting test results, discrepancies between IQ's or psychometric patterns, have proved to be at least as valuable as any single test rating, for they make multiple diagnosis possible. Thus, cases with a high performance IQ, relative to their verbal IQ, are diagnosed as possessing an intellectual capacity to at least the level of the performance IQ and a language development to the level of the verbal rating. This always holds true providing there are no other factors operating to influence the scores.

This interpretation might be rejected on the grounds that it reduces verbal intelligence tests to the status of language aptitude or achievement scales. This is not true, but in passing it is interesting to note that the authors of one of the leading text books on psychological tests,[2] maintain that for children, verbal or linguistic tests of general intelligence "are in the main, measures of 'general' scholastic aptitude, *i.e.*, of the ability to perform well the work of the school. For adults they are measures of the ability to perform such occupational tasks as depend mainly upon schooling." It is our contention that verbal intelligence tests, like all psychological tests, measure all dimensions of behavior simultaneously, but in different degrees. Verbal intelligence tests differ from other types of psychological tests in that they measure, primarily, that dimension which is largely responsible for verbal development.

Even after a verbal and performance test has been administered there is no assurance that the higher score signifies the subject's native mental capacity. In fact, there is a great deal of evidence that performance scores, too, are lowered as measures of intelligence by certain non-intellectual factors. The most important factor or dimension of behavior will be referred to, for present purposes, as mental control. The term (mental control) resembles the psychiatric concept of mental instability and is manifested on tests demanding sustained memory, concentration, accuracy, planfulness, efficiency, relevancy, and the like. All performance tests contain items which require the exercise of some or all of these attributes, hence, they are sensitive to the presence of different degrees of mental control. In 1930, Babcock[1] studied state hospital patients with two groups of psychological tests. One group contained items requiring a large degree of efficiency; the other contained those requiring a minimum efficiency. Her findings led her to advocate the use of two separate ratings—one for intelligence and one for mental efficiency. Jastak[5] compared the verbal and performance IQ's of well-behaved and poorly-behaved school children. He concluded, among other things, that "scores derived from performance tests may not be used as accurate indices of intellectual capacity if the child is known to be suffering from personality difficulties such as occur in psychosis, psychopathy, neurosis, and other abnormal states of the mind, except mental deficiency." The same investigator[4] administered the Terman Vocabulary, Stanford-Binet, and Army Performance Scale to eighty-six state hospital patients. The average Vocabulary IQ was 96.5, the average Binet IQ was 84.9, and the average Performance IQ was 73.4. All differences were statistically reliable. Uhler[7] studied the "whole child" in relation to his psychometric pattern and concluded that: "There is a very strong trend toward the development of serious personality defects and psychoses among children with wide disparities, especially when the psychomotor ability [according to the performance score] is lower. It is in this group also that the emotional instabilities and schizoid reactions predominate."

Our analysis of intelligence tests thus far follows this trend: In order to diagnose intelligence, at least one verbal and one performance intelligence test must be administered. Subjects with a language retardation will score higher on performance tests while subjects with poor mental control will score higher on verbal tests, provided no other complicating factors are operative. This seems simple and logical,

but what changes will occur in test results when a subject possesses both a language retardation and inferior mental control? From what has been discussed we would expect both scores to be lowered, each in accordance with the amount of its depressing factor. Thus a child of average intelligence with mild instability and a marked reading disability might score a Stanford-Binet IQ of 73 and a Cornell Performance IQ of 84. Or a neurotic adult of superior intelligence and a slight language retardation might make a Stanford-Binet IQ of 101 and an Army Performance IQ of 87. In both instances the higher IQ underrates the native level of intelligence, but, since the diagnosed level of intelligence is still within the average range, this underestimation is relatively unimportant. The picture changes, however, when a subject of borderline intelligence suffers a language retardation and mental instability, for in this instance a drop of one category classifies him as mentally defective. Such an error may have serious consequences from the individual and social point of view.

The remainder of this paper will be devoted to a discussion of a study made on cases whose Binet and performance scores classify them as feebleminded at the moron level. If our assumptions are valid, we should expect to find certain reliable intra-test differences between real morons and "test morons," *i.e.*, those who have moron IQ's because of the undermining influences of non-intellectual factors.

The records of eighty-three cases with Binet and performance IQ's between 50 and 70 were selected from the files of the Delaware Mental Hygiene Clinic. The test results of several cases, selected at random, were studied to learn how one subject's test responses were similar or different from another's. The most outstanding difference was found in compositions of Binet scatter. It was noted that some Binet IQ's were derived from subscores which, for the most part, varied within the limits of the feebleminded category, while others were derived from scores which varied over a wide range beginning at the defective classification and extending through to the average intelligence levels. Since interest was focused on differences in mental abilities, and not on scatter as such, only the upper limits of scatter were utilized. A technique had to be devised which would rate each subject satisfactorily according to his highest score on the Binet Scale. Because of the wide age range of the subjects (Table I) ratings had to be in terms of mental age and IQ. The following procedure satisfied both of these requirements.

The mental age and IQ of upper mental age attainment of the Binet was ascertained by adding the number of mental age months earned at the upper limit, to a newly assumed basal taken at the year immediately below it. These illustrations will clarify the method. Nina B., age seventeen years, has a Binet IQ of 50 and an Army Performance IQ of 51. Her Binet is based at year six, and scattered tests are passed through year nine where she succeeds on one item. The mental age value of this item, which is two months, is added to the assumed basal at year eight, since it is the year immediately below the level of the upper mental age attained. This yields an upper mental age of eight years two months and an IQ of 54. Thomas D., age fifty-two, has a Binet IQ of 69 and an Army Performance IQ of 61. His Binet is based at year seven, and tests are passed through year fourteen where he has three successes. Each test at the upper range has a mental age value of four months. This sum, twelve months or one year, is added to year twelve—the assumed basal immediately below—which results in an upper mental age of thirteen and an IQ of 87. This procedure is justified on the assumption that an individual could pass all subtests below his highest level of attainment, as expressed on the Binet, were it not for the lowering effect of non-intellectual factors. Scattered successes at the higher levels of the Binet are not conceived of as special abilities or aptitudes.

All IQ's are based on actual age when the subjects are below thirteen, on the Terman-Merrill[6] age reduction table when they are between thirteen and sixteen, and on a fifteen-year-age basis when they are over sixteen. Each case was placed in one of two groups depending on whether the Binet upper attainment IQ was above or below 70—the figure commonly used to differentiate the mentally

TABLE I.—GROUP *A* AND *B* COMPARED AS TO SEX, RACE AND AGE

		Group *A*	Group *B*
Sex	Male	23	23
	Female	20	17
Race	White	26	23
	Colored	17	17
Age	Range	8 to 44 years	7 to 63 years
	Average	19 years	26 years

TABLE II.—GROUP *A* AND *B* COMPARED ON THE BASIS OF IQ's

	Average IQ		
	Stanford-Binet	Performance	Highest Binet Attainment
Group *A*	58	60	64
Group *B*	65	62	80

defective from the non-defective. Subjects with speech defects, auditory impairments, and foreign language background were eliminated. There were forty-three cases in Group *A* (Binet upper IQ below 70) and forty in Group *B* (Binet upper IQ above 70). Table I shows how these groups are constituted according to sex, race, and age. The Binet, performance, and upper Binet attainment IQ's were averaged for each group. The results are recorded in Table II. Groups *A* and *B* are reasonably equated on the basis of the Binet and performance IQ's; Group *B* is superior to Group *A*, by an average of sixteen IQ points, on the basis of the Binet upper attainment IQ's. This indicates that the average individual in Group *A* can pass only those tests placed in the feebleminded categories, while the average individual in Group *B* is capable of passing Binet tests placed at the dull average level. The average case in Group *A*, when judged by his highest attainment, is still defective, but the average case in Group *B* when judged by the same criterion is not defective. Both rate mentally deficient, however, because of the averaging process inherent in this and most tests of intelligence. Let us illustrate what occurs in the case of two sixteen-year-old boys, both of whom have a mental age of 10. One's mental age is averaged from scores of 8, 9, 10, 11, 11, 11, while the other's is a mean of 6, 7, 10, 10, 11, 16. The former is incapable of coping with problems placed above year eleven, whereas the latter can succeed on an item placed at year sixteen—the usual average adult level. Surely the latter must be more intelligent than the former to accomplish this. Yet, because of the averaging technique, both are rated as possessing identical mental ages and both are diagnosed as feebleminded.

Binet and performance subtests, like verbal and performance scales, are sensitive to the extra-intellectual test depressors discussed. The higher subtests, therefore, indicate potential level of intelligence, the intermediate and lower ones denote different degrees of test lowering

due to the different multiple demands of each item. Those capable of passing a reliable subtest with a mental age value which will yield an IQ significantly above 70 can hardly be called feebleminded. Since the psychologists at the Mental Hygiene Clinic of Delaware analyze their doubtful mentally deficient cases as explained, it will be of value to note the trend of diagnosis for each group. According to intelligence ratings given in the psychologists' reports, seventy-eight per cent of the cases in Group *A* were judged to be defective and 22 per cent to be not defective. The trend is reversed for Group *B* with twenty-three per cent judged to be defective and seventy-seven per cent not defective. These percentages would be more extreme if only tests results were taken into account in diagnosing mentality.

SUMMARY

Verbal intelligence test IQ's are lowered by the presence of a general language retardation which differs from linguistic underdevelopment inherent in subjects of a foreign language background or in subjects possessing a speech or an auditory impairment. Performance test IQ's are lowered by the presence of inferior mental control. IQ's from both types of tests are depressed when language retardation and poor mental control are present in the same individual. The lowering of each IQ is dependent on the amount of each factor present. When both factors are sufficiently influential, subjects with average or above average native intelligence achieve scores in the dull average or borderline category of intelligence, while those possessing dull average or borderline native intelligence make scores that are classified as feebleminded. Because these depressing factors influence subtests of batteries as they do the batteries themselves, there is a marked difference in the range of subtests passed by the feebleminded and the pseudo-feebleminded. Specifically, the feebleminded Stanford-Binet IQ's are averaged from subtests which fall mostly within the feebleminded category when scored separately. On the other hand, the pseudo-feebleminded IQ's are averaged from subtests which fall partly within and partly above the defective category. Subscores within the defective class represent the influence of the test depressors, those above it represent the subject's potential or native intellectual capacity. Thus the real feebleminded may be distinguished from the pseudo-feebleminded by critically examining the range of subtest successes on the regular scales of intelligence.

BIBLIOGRAPHY

1. Babcock, H.: "An Experiment in the Measurement of Mental Deterioration." *Arch. Psychol.*, 1930, No. 117.
2. Garrett, H. E. and Schneck, M. R.: *Psychological Tests, Methods, and Results.* Harper & Brothers, New York, 1933.
3. Glanville, A. D.: "Psychometric Patterns of Industrial School Boys." *Delaware State Med. Jour.*, Vol. IX, 1937, pp. 91–94.
4. Jastak, J.: "Psychometric Patterns of State Hospital Patients." *Delaware State Med. Jour.*, Vol. IX, 1937, pp. 87–91.
5. Jastak, J.: *Variability of Psychometric Performances in Mental Diagnosis.* New York City, 1934.
6. Terman, L. M. and Merrill, M. A.: *Measuring Intelligence.* Houghton Mifflin Company, Boston, 1937.
7. Uhler, C.: "Wide Span Psychometric Patterns." *Delaware State Med. Jour.*, Vol. IX, 1937, pp. 65–71.

STUDENT EMPLOYMENT AND STUDENT CLASS LOAD

J. D. HEILMAN

Colorado State College of Education

A college student is advised by some of the literature on student counseling to reduce his class load by as many hours per week as he devotes to employment per day. Thus, if he works one hour per day he should carry a class load of fifteen hours; if he work two hours per day he should carry a class load of fourteen hours per week and so on. The purpose of this study was to find out whether this advice was sound from the standpoint of succeeding in college as determined by college marks.

The subjects of the study were the three hundred ninety members of the freshman class in Colorado State College of Education for the Fall quarter of 1937. The results for the sexes and for the nineteen boys who played football are reported separately. Of the girls one hundred sixty were not employed and ninety-two were employed from one-half to six hours per day. Of the boys forty-nine were not employed and seventy were employed from one to over eight hours per day. Nineteen additional boys were employed and played football. Fuller descriptions of the groups not including the football group appear in Tables I and II. In these tables are listed for the Fall quarter of 1937 the means, standard deviations, and extreme ranges of the distributions of average grades, of intelligence test scores, of the class load in hours per week, and of the hours of employment per day.

The intelligence of the groups was measured by means of the 1937 form of the American Council on Education Psychological Examination. The results are expressed in terms of PE scores. The marks are expressed in terms of a five-point system in which one denotes a failing mark, three an average mark, and five the highest mark. Two is used to designate a mark midway between one and three, and four a mark midway between three and five. The normal class load in Colorado State College of Education is seventeen hours per week for the freshman student. This load includes one hour for physical exercise courses.

According to the means reported in Table I the group of employed girls makes a higher grade average than the group of unemployed girls in spite of the fact that the employed group devotes an average of three hours per day to employment. However, in intelligence the

employed group has the higher mean, and in class load it has the lower mean. The difference between the class-load means is only .63 of an hour. This difference was regarded as negligible. The difference between the means for intelligence is .24 of a PE score or .16 of an SD score. If the mean of the average grades for the unemployed group

TABLE I.—MEANS, STANDARD DEVIATIONS, AND EXTREME RANGES OF THE AVERAGE GRADES, INTELLIGENCE TEST SCORES, CLASS LOAD IN HOURS PER WEEK, AND EMPLOYMENT IN HOURS PER DAY FOR NINETY-TWO FRESHMAN GIRLS WHO WERE EMPLOYED AND, EXCEPTING EMPLOYMENT, FOR ONE HUNDRED SIXTY FRESHMAN GIRLS WHO WERE NOT EMPLOYED

	Mean		Standard deviation		Extreme range	
	Em-ployed	Unem-ployed	Em-ployed	Unem-ployed	Em-ployed	Unem-ployed
Average grades.....	3.18	3.07	.72	.83	1.7– 5.0	1.0– 5.0
Intelligence.....	7.72	7.48	1.38	1.49	4.6–11.2	0.0–11.0
Class load.....	16.24	16.87	1.36	1.05	12.0–18.0	12.0–22.0
Employment.....	3.02	.....	1.18	....	0.5– 6.0	.....

TABLE II.—MEANS, STANDARD DEVIATIONS, AND EXTREME RANGES OF THE AVERAGE GRADES, INTELLIGENCE TEST SCORES, CLASS LOAD IN HOURS PER WEEK, AND EMPLOYMENT IN HOURS PER DAY FOR A GROUP OF SEVENTY FRESHMAN BOYS WHO WERE EMPLOYED, AND, EXCEPTING EMPLOYMENT, FOR A GROUP OF FORTY-NINE FRESHMAN BOYS WHO WERE NOT EMPLOYED

	Mean		Standard deviation		Extreme range	
	Em-ployed	Unem-ployed	Em-ployed	Unem-ployed	Em-ployed	Unem-ployed
Average grades.....	2.86	2.58	.83	.79	1.0– 4.7	1.0– 5.0
Intelligence.....	7.54	7.05	1.57	1.23	3.9–11.2	4.5–11.0
Class load.....	15.73	16.57	1.91	1.99	10.0–18.0	12.0–21.0
Employment.....	3.20	.....	1.72	....	1.0– 8.+	.....

be raised .16 of the SD of the distribution of the average grades, it will be 3.20. The difference between this mean and the corresponding mean for the employed group is negligible. It may, therefore, be concluded that the group of employed girls has an academic standing which is equal to that of the group of unemployed girls.

The employed boys, as shown in Table II, also make a higher grade average than the unemployed boys in spite of the fact that they devote more than three hours per day to employment. As is the case with the girls the mean intelligence test score is higher and the mean class load is lower for the employed than for the unemployed group. The difference between the means of the class loads is only .84 of an hour. The difference between the means of intelligence is .49 of a PE score or .33 of an SD score. If the mean of the average grades for the unemployed groups be raised .33 of the SD of the distribution of the average grades, it will be 2.84, which is .02 less than the corresponding mean for the employed group. Therefore, the academic standing of the employed group is just as high as that for the unemployed group.

Out of the total number of one hundred thirty-eight freshman boys, nineteen were employed and also were members of the freshman football team. For this group the mean of the average grades for the Fall quarter was only 2.22. This is .54 of a step below the corresponding mean for the seventy freshmen who were employed but did not play football. As the standard error of this difference is .21, the difference comes very close to being a reliable one. However, other factors must be considered in interpreting this mean. In comparison with the boys who were employed but did not play football, those who worked and played football spent about the same number of hours in employment, carried a class load which was smaller by 1.47 hours per week, and had a mean intelligence test score which was lower by .66 of a PE score. When their grade mean was corrected to allow for this difference in intelligence it was only .27 of a step below that of the other group. It should also be pointed out that the football group devoted from three to four hours per day to football. This not only required much time but was also very fatiguing.

According to the figures listed after employment in Tables I and II, the hours of employment per day vary for the girls from one-half to six and for the boys from one to over eight. However, only eight of the ninety-two girls worked more than four hours per day. Because the members of each of the two working groups vary considerably in the amount of time devoted to work it is possible to determine the effect of the amount of work upon the excellence of the marks by means of the correlation technique. Inter-correlations and semi-correlations among the four variables of marks, intelligence, class load, and employment were computed. The results of these computations are presented in Tables III and IV.

By an examination of the zero order correlations of Tables III and IV, it may be seen that college marks yield a marked correlation with only one of the other three variables. This variable is intelligence The coefficient for marks and employment is slightly positive for th

TABLE III.—INTER- AND SEMI-PARTIAL CORRELATIONS AMONG THE GRADES INTELLIGENCE TEST SCORES, STUDENT EMPLOYMENT, AND CLASS LOAD OF NINETY-TWO FRESHMAN GIRLS

	Intelligence (I)	Employment (E)	Class Load (L)
Average grade (G).....	.592 ±.045	.148 ±.069	.041 ±.070
Intelligence.....	....	−.126 ±.069	.027 ±.071
Employment.....	....	....	−.442 ±.057

$r_{(G.I)E} = .276 \pm .065$ $r_{(G.LI)(E.L)} = .323 \pm .063$

TABLE IV.—INTER- AND SEMI-PARTIAL CORRELATIONS AMONG THE GRADES, INTELLIGENCE TEST SCORES, STUDENT EMPLOYMENT, AND CLASS LOAD OF SEVENTY FRESHMAN BOYS

	Intelligence (I)	Employment (E)	Class Load (L)
Average grade (G).....	.560 ±.055	−.146 ±.079	.071 ±.080
Intelligence.....	....	.0004 ±.081	−.005 ±.081
Employment.....	....	....	−.085 ±.080

$r_{(G.I)E} = -.176 \pm .078$ $r_{(G.LI)(E.L)} = -.169 \pm .079$

girls and slightly negative for the boys. If employment had any significant effect upon marks the coefficients should have been markedly negative. The negligible amount of correlation between marks and employment can not be attributed to a higher degree of intelligence by those who worked the more hours, because the coefficient between

intelligence and employment is slightly negative for the girls and zero for the boys. If the students who were employed the more hours would have had the higher intelligence these coefficients should have been positive. However, negative coefficients were obtained for employment and class load. This means that the students who devoted the more hours to employment carried the lighter class load. Perhaps if the class loads had been equal for the members of each group, the coefficient between grades and employment would have been negative.

To determine the coefficient of correlation between grades and employment freed from the influence of the disturbing factors of differences in intelligence and in class load semi-partial coefficients were computed. These appear at the foot of each of the last two tables. Two semi-partials were computed, $r_{(G.I)E}$ and $r_{(G.IL)(E.L)}$. The formula for $r_{(G.I)E}$ yields the coefficient between the hours of employment and college marks freed from the effect of differences in intelligence. This coefficient is .28. As the corresponding zero order coefficient is only .15 the evidence for the improvement of grades with an increase in the amount of work up to four hours per day has been strengthened by allowing for the disturbing factor of intelligence.

The formula for $r_{(G.IL)(E.L)}$ yields the coefficient between the hours of employment freed from the effect of differences in class load with the college marks freed from the effect of differences in intelligence and class load. This coefficient is .32. Therefore, it appears that when allowance has been made for both of the disturbing factors of intelligence and class load, the tendency for marks to improve with an increase in the hours of employment is still stronger. It is not improbable that the student who is obliged to earn part of the expenses of a college education may take his college work more seriously than the student who does not have to work and consequently tends to do better in spite of his employment.

For the group of seventy boys the zero order coefficient between grades and employment is $-.15$; the semi-partial $r_{(G.I)E}$ is $-.18$; and the semi-partial $r_{(G.IL)(E.L)}$ is $-.17$. These low negative coefficients are far too small to be statistically reliable. It is, therefore, impossible to tell whether for the boys there is a slight tendency for grades to be reduced by the amount of employment. The fact that the girls show a stronger tendency in the opposite direction might be attributed to a larger variation in the boys' hours of work and to more fatiguing work on the part of the boys.

There is at least one more factor, that of health, which should have been taken into consideration in our investigation. Because data on the effect of employment upon the health of our students were not available they were not included in our study. It is, however, not likely, excepting possibly for a few cases, that the health of our employed students is impaired by their work, because the average freshman devoted only fifteen hours per week to his studies. With no employment he could devote from thirty-five to forty hours per week to his studies. He, therefore, has from twenty to twenty-five hours per week to devote to employment. Healthy students who have the time to and do sleep eight hours per day will not have their health impaired by the amount of healthy study and work which our students engage in.

The findings of this investigation do not justify the establishment and enforcement of rules for the purpose of restricting the class load of a college student on account of employment within the limits of four hours per day. Doubtless the health of a small percentage of students is impaired by an excessive load of studies, employment, and extra-curricular and social activities. I believe that the activities of such cases should be regulated by the health officers. Moreover, it should be remembered that there are values in life other than health, and that it is frequently desirable to sacrifice some health, especially if of a transitory nature, for those other values.

It should also be pointed out that the findings for Colorado State College of Education may not hold for all other colleges. It is probable that in some colleges the freshman student may be motivated to devote more than fifteen hours per week to his studies. If it were necessary for him to spend forty hours per week on his studies, employment might lower his college marks considerably.

THE USE OF VISUAL, AUDITORY, AND KINESTHETIC IMAGERY IN THE TRANSFER OF MUSICAL NOTATION TO THE PIANO KEYBOARD

KENNETH L. BEAN

Baton Rouge, La.

The question of the degree to which any one type of imagery guides the musician in reading new music at the piano is one which cannot be settled by the performer's introspections. Some pianists believe that their visual impression of the music is transferred into an auditory image of the way that it should sound when played, and that this auditory image, in turn, guides their fingers. Others describe their procedure as a direct mechanical response to the visual note. They think of it in terms of action on the keyboard without reference to sound as far as they are aware of the process.

To what extent these individuals are clearly aware of what they do in the situation of playing at sight is not known. Systems of well-fixed habits which make up this highly complex skill may take care of this situation largely subconsciously in experienced performers. Therefore, I have attempted to test out each kind of imagery objectively in a situation requiring its use. By comparison of performance with various types of imagery on the same kind of musical material, I can arrive at a fair estimate of the accuracy with which each player can employ any given type as a guide in his performance. This does not mean, however, that he actually makes most use of the sort of imagery which he has most clearly.

The individual differences in imagery do not receive attention in the few studies of music reading which have been published, since these scientific investigations deal almost exclusively with the perceptual phase of the process. Seashore has discussed musical imagery, but not as it is applied to reading. The clear auditory imagery possessed by composers has been discussed by many writers to great length, and the visual memory of a page of music has been demonstrated repeatedly by many a performer and teacher. The latter phenomenon has been known to enable many performers to continue playing after his auditory memory had temporarily failed him in the middle of a composition.

The only study known to me which is closely related to the present problem is that of Murray, in an unpublished thesis, *How to Memorize Music.*[3] The problem was a rather vague and general one, and the

experiments were none too well controlled. However, the subjects were given selections to memorize visually on which they were tested. Selections of equal difficulty were presented for auditory memory under otherwise similar conditions. Identification of the passage learned when presented with other selections was used as a test in both cases. Kinesthetic memory was tested in a similar way on a keyboard producing no sound. The series of pressures was conveyed to the subject mechanically through the keys hidden from his sight. Combinations of visual-auditory and visual-motor memorizing were also attempted. The conclusions drawn by Mrs. Murray do not seem particularly significant except that there were pronounced individual differences in imagery dominance which determined the method which worked best for each subject.

In my recent experimental work on span of perception in music reading, there occurred a number of types of errors. Usually a given individual was consistent in having one kind of difficulty which showed itself in the errors that he made. This observation in a number of cases indicated difficulty in transfer to the keyboard for reasons not clearly understood. Some cases of vague and short-lived visual impressions were evident. Out of these interesting possibilities, I developed the method of this study.

PROCEDURE

The experiment was divided into five steps.

(1) The Seashore Tonal Memory and Rhythm tests were given to determine the clearness of these auditory impressions.

(2) The same material as in the two Seashore tests was placed on cards in the form of musical notation. The cards were then presented in mixed order visually, the task of the subject being exactly as in the first case except that the subjects read through once each visual pattern instead of listening to the corresponding auditory pattern. A double-exposure tochistoscope regulated the time of each exposure, which was approximately that taken to play the individual pattern on the record. Time between exposures to be compared was also made equal to the interval between the corresponding patterns on the Seashore record. One hundred responses were recorded for each subject on each test. This step in the experiment was for comparison of the clearness of visual and auditory impressions of the same musical material.

(3) Successions of tones similar to, but not identical with, Seashore's Tonal Memory patterns were played on the violin. The object in using this instrument was to match the sustained tones of Seashore's record as nearly as possible, since piano tones fade rapidly and differ greatly from those on the record in their quality. Simultaneously with this auditory presentation from behind a screen, a card was exposed to the subject, on which he read a series of notes. He was told that these notes might be exactly like the tones he was hearing, or that one tone might be different from the correct pitch of the note written on the card. In other words, he might be looking at *C, F, E,* and hearing *C♯, F, E.* As in Seashore's Tonal Memory test, there was never more than one note changed, and it was either raised or lowered a half step. In two cases the visual and auditory patterns were identical. The subject indicated all changes by giving the number of the note in the series that was altered, as in Seashore's Tonal Memory test.

After this procedure was completed, a similar one was followed with rhythms closely resembling Seashore's patterns. They were played on a telegraph key, while the observer read these, or slightly changed patterns, in the form of notation concealed behind a screen. The responses were written simply as *same* or *different,* (*S* or *D*). The object of this step was to determine how accurately the reader could hear mentally what he saw in notation.

(4) The subject was seated before a piano keyboard of the sort employed in teaching keyboard harmony, on which no sound could be produced, but the keys respond to pressure just as regular piano keys do. He was then told that he would hear a successsion of tones, all within the octave above middle *C.* In response to each tone heard, he was instructed to play the correct note immediately if he could. Since none of the subjects had so-called absolute pitch, the starting note of each pattern had to be named by the experimenter before the pattern was played. Once located definitely on the keyboard, the subject should logically be able, by perception of the intervals between the tones, to follow through each of the simple melodic patterns with but little difficulty if his auditory imagery guides him at all accurately in playing new material at sight. The nature of the errors was determined as far as possible by observation.

As a corresponding task for rhythmic material, the subject was given auditory patterms tapped by the experimenter on a hidden telegraph key, to which he was to respond by reproducing them on

experiments were none too well controlled. However, the subjects were given selections to memorize visually on which they were tested. Selections of equal difficulty were presented for auditory memory under otherwise similar conditions. Identification of the passage learned when presented with other selections was used as a test in both cases. Kinesthetic memory was tested in a similar way on a keyboard producing no sound. The series of pressures was conveyed to the subject mechanically through the keys hidden from his sight. Combinations of visual-auditory and visual-motor memorizing were also attempted. The conclusions drawn by Mrs. Murray do not seem particularly significant except that there were pronounced individual differences in imagery dominance which determined the method which worked best for each subject.

In my recent experimental work on span of perception in music reading, there occurred a number of types of errors. Usually a given individual was consistent in having one kind of difficulty which showed itself in the errors that he made. This observation in a number of cases indicated difficulty in transfer to the keyboard for reasons not clearly understood. Some cases of vague and short-lived visual impressions were evident. Out of these interesting possibilities, I developed the method of this study.

PROCEDURE

The experiment was divided into five steps.

(1) The Seashore Tonal Memory and Rhythm tests were given to determine the clearness of these auditory impressions.

(2) The same material as in the two Seashore tests was placed on cards in the form of musical notation. The cards were then presented in mixed order visually, the task of the subject being exactly as in the first case except that the subjects read through once each visual pattern instead of listening to the corresponding auditory pattern. A double-exposure tochistoscope regulated the time of each exposure, which was approximately that taken to play the individual pattern on the record. Time between exposures to be compared was also made equal to the interval between the corresponding patterns on the Seashore record. One hundred responses were recorded for each subject on each test. This step in the experiment was for comparison of the clearness of visual and auditory impressions of the same musical material.

(3) Successions of tones similar to, but not identical with, Seashore's Tonal Memory patterns were played on the violin. The object in using this instrument was to match the sustained tones of Seashore's record as nearly as possible, since piano tones fade rapidly and differ greatly from those on the record in their quality. Simultaneously with this auditory presentation from behind a screen, a card was exposed to the subject, on which he read a series of notes. He was told that these notes might be exactly like the tones he was hearing, or that one tone might be different from the correct pitch of the note written on the card. In other words, he might be looking at *C, F, E,* and hearing *C#, F, E.* As in Seashore's Tonal Memory test, there was never more than one note changed, and it was either raised or lowered a half step. In two cases the visual and auditory patterns were identical. The subject indicated all changes by giving the number of the note in the series that was altered, as in Seashore's Tonal Memory test.

After this procedure was completed, a similar one was followed with rhythms closely resembling Seashore's patterns. They were played on a telegraph key, while the observer read these, or slightly changed patterns, in the form of notation concealed behind a screen. The responses were written simply as *same* or *different,* (*S* or *D*). The object of this step was to determine how accurately the reader could hear mentally what he saw in notation.

(4) The subject was seated before a piano keyboard of the sort employed in teaching keyboard harmony, on which no sound could be produced, but the keys respond to pressure just as regular piano keys do. He was then told that he would hear a successsion of tones, all within the octave above middle *C.* In response to each tone heard, he was instructed to play the correct note immediately if he could. Since none of the subjects had so-called absolute pitch, the starting note of each pattern had to be named by the experimenter before the pattern was played. Once located definitely on the keyboard, the subject should logically be able, by perception of the intervals between the tones, to follow through each of the simple melodic patterns with but little difficulty if his auditory imagery guides him at all accurately in playing new material at sight. The nature of the errors was determined as far as possible by observation.

As a corresponding task for rhythmic material, the subject was given auditory patterms tapped by the experimenter on a hidden telegraph key, to which he was to respond by reproducing them on

another key. Careful introspections were recorded on this activity. The nature of errors was observed as closely as possible.

(5) The same material that was used in the preceeding part of the experiment was exposed on cards in the form of notation, the time being controlled so as to permit the subject to read each pattern once. He played the notes in tonal series on the silent keyboard, and played the rhythms on the telegraph key. Errors were recorded and studied closely. Introspections were used with the objective data.

The ten subjects were college students. Seven of them were engaged in advanced study in the Louisiana State University School of Music. Three of these were outstanding performers. Of the three outside the School of Music, one was a pianist of unusual ability who had taken a thorough conservatory course. The other two were less advanced in music. All had some technical knowledge of the piano.

RESULTS

The data obtained in terms of per cent right are shown in Table I.

TABLE I

Subject	Tonal memory	Rhythm	Visual tonal memory	Visual rhythm	Tones visual and auditory	Rhythms visual and auditory	Playing heard tones	Playing heard rhythms	Playing seen	
									Notes	Rhythms
1	91	83	94	98	40	78	57	95	99	81
2	85	91	100	98	98	92	22	77	100	97
3	87	71	94	90	42	88	18	45	100	91
4	72	76	96	98	31	77	11	36	100	94
5	83	73	98	96	50	84	9	50	100	97
6	72	87	96	96	45	76	0	58	100	95
7	98	94	98	100	44	75	20	88	100	88
8	78	84	98	94	36	56	2	42	95	50
9	97	85	98	94	62	88	12	72	93	82
10	88	75	86	92	52	58	4	42	76	42
Mean...	85.1	81.9	95.8	95.6	50	77.2	15.5	60.5	96.3	81.7

The subjects are listed here in order from the most highly trained to the least trained in music. From these figures the fact is clear that, in general, visual stimuli gave the subjects clearer impressions resulting in more accurate comparisons and better transfer into action than did auditory presentation of the same musical material. Par-

ticularly was this evident in Parts 4 and 5, in which actual playing was done by the subjects. Comparison of the visual with the auditory presentation of the Tonal Memory material reveals that the visual scores were higher in eight cases; in one they were equal; and in the last case the auditory score was higher, though not significantly so, probably because of this subject's lack of experience with musical notation.

With rhythms the advantage of visual presentation was more pronounced. In every case visual comparisons were better than auditory, though for some individuals this difference was relatively small. Introspections in six cases showed that the observers could get whole patterns more easily by seeing them than by hearing them. These individuals reported that the visual presentation was ordinarily grasped all at once, while auditory material was a succession of stimuli as a rule, which made it difficult to anticipate what might follow after hearing the first of each melody or rhythm. The nature of errors made in reproducing material given in these two ways indicates that the observers were right in this interpretation. Certainly the impression of the whole seemed to be more definite in most cases when the stimulus was visual. Subject (1) was an exception to this rule, however. He could play more accurately a rhythm which he heard, but with tones he was lost as to absolute location on the keyboard when he heard them, and, like the rest, played them better at sight.

Introspective reports revealed that the auditory patterns in the comparisons were not transferred into visual impressions at all, but two observers[3,4] reported that they transferred the patterns into incipient action in the fingers, especially with rhythms. The experimenter observed these motor tendencies occasionally in these two and also, to a lesser degree, in subject (8), who was not aware of having moved her fingers in response to the tonal series or rhythms.

The visual patterns to be compared in Part 2 were all perceived as visual material, and were not heard mentally except by subject (1), who tended to hear the rhythms in spite of his efforts to compare them visually, which the experimenter encouraged all subjects to do.

Part 3 was a source of difficulty for most of the readers. Even some with thorough theoretical as well as practical training were unable to hear mentally with much accuracy what they saw. This was especially true with tonal series. Rhythms were easier to transfer. In the case of subject (4), the rhythms caused some motor response of the hand, which she compared with what she heard. Courses in

theory and ear-training enabled some of the observers to identify the intervals between the tones heard in series, and compare them with the intervals as written. None of the subjects had absolute pitch; therefore, exact auditory imagery of the sound of each note was impossible. Rhythms could be imaged more accurately as to details. Lack of training lowered the scores of subjects near the end of the table, while the extremely high scores of (2) on this task may be accounted for by the fact that she is a violinist, and violinists must listen carefully to every detail of their performance in order to be at all successful.

From the data obtained in Parts 4 and 5, one can find no evidence that these ten musicians are able to play notes on the piano accurately after receiving an auditory impression. Some can reproduce heard rhythms with few errors, but the majority had difficulty with this task also. The visual presentation, however, was an easier method for the tonal series than it was for the rhythmic patterns, though in both kinds of material the subjects far excelled their performance in Part 4.

Introspections showed that in all but two cases the observers believed that they thought of the notes on the cards in terms of the pressing of keys, (action), not stopping to imagine how each note would sound on the piano. Subject (1) insisted that he could not play anything right until he had an idea how it sounds when played, yet even he, when presented with the actual sounds in the tonal series, made a surprisingly large number of errors compared with those that he made when reading the same material at sight. He scored better on rhythms when he heard them, however.

A brief review of individual difficulties as indicated by the specific errors made by each subject on each pattern seems appropriate here. Subject (1) is a gifted pianist with a thorough knowledge of theory who has played a broad repertoire of very difficult music from memory. He has indeed a phenomenal memory for music, but reads it slowly with a great deal of stumbling in spite of his persistent efforts, under the guidance of able teachers, to overcome this fault. He served as subject in my earlier experiment,* and was unable to gain much in reading efficiency even though his span of perception improved from one note at a time to slightly below the average of the practice group. He has a decided auditory dominance. His difficulty seems to lie in transfer of the visual impression into an auditory image or into

* See Bibliography No. 1.

action. His reaction time averages about two seconds for all cards to be read, but is a small fraction of a second when he plays from an auditory impression.

Subject (2) is a gifted and accomplished young violinist who has played a wide repertoire of the most difficulty level from memory. She also reads well for either piano or violin. Her scores are uniformly high except for the tonal series in Part 4, in which she could not get the exact location of the notes on the keyboard, even though she had a fair idea of every pattern.

Subject (3), a talented, advanced piano student, has fair ability in reading music. Visual imagery seems dominant in these test situations. She had difficulty with anything auditory having to do with tones except the Tonal Memory test. She could image rhythms in terms of sound fairly well.

Subject (4) is advanced technically on the piano and reads notes well, but memorizes with great difficulty. She is not considered talented by her teacher. She has clear visual imagery, but very little auditory imagery. Consequently she did better with anything she saw than with anything she heard. She reported that she thought of all notes in terms of the act of playing them, or muscular movement. If she tried any other scheme, she failed miserably. Sound seemed to guide her very little.

Subjects (5) and (6) were similar to (4) except less advanced in piano technique and less inclined to think of musical notations in terms of muscular actions involved in playing.

Subject (7) is a gifted pianist who shows evidence of real musicianship. She plays a fairly large repertoire from memory, memorizes easily, but reads slowly with much stumbling. Her scores indicate good visual and auditory imagery, but difficulty in transfer of visual stimuli into auditory imagery, particularly with tones. She was able to play from an auditory stimulus better than the average, however, and the cause of her reading difficulty does not come out in the errors which she made in this experiment.

The last three observers listed in the table were not advanced in music, and lack of familiarity with notation was evident in Parts 3 and 5. Subject (9) appears to have talent, but his lack of theoretical knowledge made him unable to identify tonal intervals he heard and relate them to the keyboard. He also failed to recognize common rhythmic figures in notation. Subject (10), a beginner in music, had a marked defect in motor rhythm.

A few additional trends in the figures deserve mention. Scores on the two Seashore tests in Part 1 are dependent upon innate capacities as the author of the tests has shown, but the ability to transfer this same or similar material into action on a piano keyboard or telegraph key seems to depend largely upon various factors in the training of each individual. If you have responded to the sound of, or sight of, middle C by pressing the appropriate key so many times that it has become well-fixed habit, then you can easily think of the note in terms of the act. The more experienced pianists could transfer a stimulus into such action more quickly and with greater accuracy than could the less trained. Introspections revealed that some subjects thought of "how it would feel to play the music," rather than how it would sound. The individual differences in the use of these types of imagery seemed to be differences in relative clearness and frequency of appearance. In no case was one class of imagery completely absent.

CONCLUSIONS

This research has only attempted to survey the possibilities of study of the use of imagery in music reading. Since the limited number of subjects does not make a fair sample for statistical purposes, the emphasis has been placed on individual differences here. However, since the experimental situations tried have proved fruitful, they may be repeated with larger groups of music students in order that more definite conclusions may be reached.

Within the limits of the ten individuals tested, the following conclusions seem evident:

(*a*) For most of the subjects visual imagery was clearest and was used most in the experiments.

(*b*) None of the readers could translate seen notes into mentally heard tones which, in turn, accurately guided their performance, but some were vaguely aided by the way in which they thought the notes should sound.

(*c*) The best readers had high scores on all the tasks, except possibly the tonal series in Part 4, which seems not to be essential to good reading. A low degree of both visual or auditory imagery seems to have resulted in reading disability.

(*d*) For reading situations a note on the page means to most performers the act of pressing a key, not a sound. Rhythmic figures

are also usually interpreted in terms of action, but a few readers hear them when they look at them.

Probably further testing will result in the discovery of some musicians who are more capable of handling all these tasks than the ten reported in this study, and, doubtless, others who are less capable, but at least these interesting cases have been given a fairly complete analysis. We may logically assume that if an individual cannot perform accurately these tasks requiring the use of each kind of imagery, he is aided very little in his reading by the types of imagery in which he is weak. The results of this investigation do not reveal with certainty that he uses his clearest type of imagery most in the practical reading situation, but they do demonstrate the degree of accuracy possible for him with each type.

BIBLIOGRAPHY

1. Bean, K. L.: "An Experimental Approach to the Reading of Music." *Psychol. Monog.*, 1938, Vol. L, No. 6.
2. Jacobsen, Ole: *An Analysis of Eye-Movements in Reading Music*, Thesis, University of Chicago, 1926.
3. Murray, Fay Tyler: *How to Memorize Music*, Thesis, (unpublished), Louisiana State University, 1925.
4. Ortmann, O.: "Elements of Chord Reading in Musical Notation." *J. Exper. Educ.*, 1934, Vol. III, pp. 50–57.
5. Ortmann, O.: "Span of Vision in Note Reading." *Yearbook of the Music Education National Conference*, 1937, pp. 88–93.
6. Wever, Homer: *An Experimental Study of Music Reading*, Thesis, Stanford University, 1930.

TWINNING AS A FACTOR INFLUENCING PERSONALITY

LILLIAN PORTENIER

University of Wyoming

Many psychological studies of twins have centered about the question of the influence of heredity and environment, or nature *vs.* nurture. The results of most of these studies have been interpreted to indicate that physical characteristics are least affected by environment, that intelligence is affected more, educational achievement still more, and personality or temperament as measured by the tests used the most. Studies using subjects other than twins tend to verify the conclusion that heredity is not a major factor in determining most characteristics of personality.

The present study is an attempt to determine how twinning as a social, not biological, factor influences personality. More specifically, the question which has been raised is: "How do twins compare in personality traits as measured by tests with ordinary siblings?" The study was suggested by a number of university instructors and high school teachers who were confronted with problems presented by one or both of a pair of twins in adjusting to various phases of campus life.

In securing data to throw light on this question, twelve pairs of twins were matched with twelve pairs of siblings who were of the same sex and approximately the same age and social status. Data were secured for five pairs of female twins, four pairs of male twins, and three pairs of mixed twins. The twins ranged in age from fourteen to twenty-three years with a mean chronological age of eighteen years. Nine pairs of the twins were enrolled for college work and three pairs were in high school. As judged by the criteria of Newman, Freeman and Holzinger[3] only two pairs of the twins were identical.

As a pair of twins was found for study, the registrar's records were examined to find a pair of sisters, or brothers, or a sister and a brother, as was the case with the twins. In all cases one sib was found who was exactly the same age or not more than one year younger than the pair of twins and the other sib not more than a year or two older than the twins. The twelve pairs of sibs ranged in age from thirteen to twenty-four years with a mean chronological age of eighteen years which was also the mean age for the twins.

A modified form of the Sims Scale was used in obtaining a score for socio-economic status of the subjects. Fifteen items identical or closely

similar to items on the Sims Scale were selected from the personal data sheet accompanying the Ohio Test as explained in a recent study by the writer, "Aspects of Socio-Economic Status Which Differentiate University Students."[4]

The following battery of eight tests was used in securing data on certain aspects of personality: The Ohio State University Intelligence Test, the Otis Self-Administering Test of Mental Ability, the Bernreuter Personality Inventory, The Maller Character Sketches, Parts I and II, the Allport A-S Reaction Study, the Strong Vocational Interest Blank, the Meier-Seashore Art Judgment Test, and the North Carolina Rating Scale. The Allport A-S Reaction Study was used merely as a check on the reliability of the scale for dominance-submission as measured by the Bernreuter Test. This and other checks later dis-

INTRACLASS CORRELATIONS, MEANS OF THE DIFFERENCES OF RAW SCORES AND MEAN PERCENTILE RATINGS FOR THE RESULTS OF VARIOUS MEASURES OF PERSONALITY

	r		s		CR	Mean difference of scores		Mean percentile rating	
	Twins	Sibs	Twins	Sibs		Twins	Sibs	Twins	Sibs
Height	.83	.39	1.19	.41	1.66	2.1	2.4	5'-4"	5'-6"[1]
Weight	.84	.40	1.23	.42	1.72	9.3	11.9	121.3	125.9[1]
OSU Test	.67	.39	.81	.41	.85	22.8	26.3	53	52
Otis Test	.73	.64	.93	.76	.36	6.3	7.2	47	47
Bernreuter Personality Inventory									
B-1 Neurotic Tendencies	.27	.54	.28	.60	.68	88.9	61.0	56	53
B-2 Self Sufficiency	.67	.21	.81	.21	1.28	32.0	55.9	33	34
B-3 Introversion-extroversion	−.02	.52	−.02	.58	1.28	66.4	35.7	43	44
B-4 Dominance-submission	.42	.58	.45	.66	.45	44.3	40.8	40	71
B-5 Confidence in Self	.21	.59	.21	.68	1.00	98.1	65.4	39	56
B-6 Sociability	.16	.53	.16	.59	.92	62.2	48.0	43	42
Maller Character Sketches									
1. Habit Patterns	.55	.23	.62	.23	.83	4.8	5.8	51	53
2. Self Control	.08	.18	.08	.18	.21	7.7	7.8	49	49
3. Social Adjustment	−.03	.71	−.03	.89	1.96	9.5	5.5	50	50
4. Personal Adjustment	.38	.38	.40	.40	.00	7.3	7.2	50	52
5. Mental Health	−.07	.01	−.07	.01	.17	5.2	1.0	39	50
6. Readiness to Confide	.37	.51	.39	.56	.36	1.8	1.5	62	62
7. Total Score	.04	.37	.04	.39	.74	24.7	22.8	52	54
Allport A-S Reaction Study	.33	.50	.34	.55	.45	15.7	13.3	41	55
Strong Vocational Interest Blank Masculinity-feminity	.92	.67	1.59	.81	1.66	22.8	28.4		
Meier-Seashore Art Judgment Test	.51	.56	.56	.63	.15	7.1	7.5	74	72

[1] Mean measures.

cussed were considered advisable, since it was realized at the outset that it would be difficult to secure a sufficiently large number of twins for statistical reliability.

Since data were secured for only twelve pair of subjects in each group, Fisher's[5] small sample formulas were used in making the comparisons. Where the tests resulted in negative as well as positive scores—as was true for the Bernreuter, Allport and Strong Scales,—the numbers were treated algebraically, with due regard to signs. Raw scores were used for the statistical treatment of results for all scales except for the Ohio Test. All subjects did not take the same form of this test, consequently, it was necessary to convert the scores into percentiles to yield comparable results. Comparisons between the two groups are based on the Fisher Intraclass Correlations for the twins and for the sibs, on the means of the differences of the raw scores for the twelve pairs of twins and for the twelve pairs of sibs, and on the mean percentile scores of the two groups.

The scores for similar traits as measured by different scales, while not identical, were consistent for the two groups as is shown by the table. That is, the correlations for the results of the Ohio Intelligence Test for the twins approximate the correlation for their results on the Otis Test. The correlations on both tests were consistently higher for the twins than for the sibs. This finding agrees with the generally established fact that twins are more similar in intellectual capacity than are ordinary sibs. Likewise, similar intraclass correlations were found for the twins and for the sibs on the scale for dominance-submission in the Bernreuter Inventory and on the raw scores of the Allport A-S Reaction Study, and for Sociability as measured by the Bernreuter Inventory, and Social Adjustment as measured by the Maller Character Sketches. The correlations for both of these tests were consistently lower for the twins than for the sibs.

No consistent differences were found as between male, female, and mixed twins with regard to sex nor between identical twins when compared with fraternal twins of approximately the same age and social status.

In determining the significance of the difference between two correlations the values of r were transformed to corresponding values of z from Table V as given by Fisher[5] and the difference in the z values divided by the standard error. According to Fisher, the difference in the z values must exceed twice the standard error to be judged significant for twelve cases. None of the critical ratios exceed twice

the standard error. A few, however, approach Fisher's criterion, since they range from once to twice the size of the standard error. Using a difference greater than the standard error, that is, a critical ratio greater than 1.00 as the criterion, aspects of personality on which the twins appear to differ more than do the sibs are introversion-extroversion, confidence in self, and social adjustment. Lower correlations, also, were found for the twins than for the sibs on neurotic tendencies, dominance-submission, readiness to confide, and general adjustment as expressed by the total score on the Maller Character Sketches.

No closely similar studies have been found with which to compare these findings. Carter[2] studied one hundred thirty-three pairs of twins classified as to type using the Bernreuter scale. The identical twins showed much greater similarity in neurotic tendencies, self-sufficiency and dominance-submission than either the like or the unlike sexed fraternal twins but little difference in the tendency toward extroversion-introversion. However, his correlations for all types of twins on the four traits are similar to the correlations found in this study.

Higher correlations for the twin pairs than for the sib pairs were found for the measures of physical traits (height and weight), intelligence, self-sufficiency, habit patterns, and masculinity-femininity as measured by the Strong Vocational Interest Test. The means of the differences in the raw scores tend to agree, on the whole, with the results of the intraclass correlations as might be expected. That is, where higher correlations were found for the twins than for the sibs the mean differences in raw scores were lower for the twins and *vice versa*.

Each of the twins was asked to state a first and a second choice of vocations from the list for which keys are available for the Strong Vocational Interest Blank. The blank for each of the forty-eight subjects was scored for at least two vocations, his own first choice and for the first choice of his twin or sib. In case of agreement in first choice for any pair of twins or sibs the blanks were scored also for the second choice of each pair. No consistent tendency was found in either the intraclass correlations or in the means of the difference of the raw scores for the results of the Strong Test.

Carter[1] found a correlation .50 for identical twins and .28 for fraternal twins in determining the similarity of occupational interests using the Strong Scale.

Again, the difference in the correlations and in the means of the differences of the scores on the Meier-Seashore Art Judgment Test were not significant.

The North Carolina Rating Scale was used to secure (1) a self-rating for each of the subjects, (2) a rating by his twin or sib, and (3) a rating by a friend. In the last instance the rating was made by a friend who had had about the same opportunity for acquaintance with both subjects of a pair of twins or sibs. The ratings were treated numerically. Comparisons were made (1) between rating of self and rating by twin or sib, (2) between rating of self and rating by the friend, (3) between ratings by twin or sib and rating by friend and, finally, (4) between the friend's rating of each pair of sibs. No consistent differences were found between the rating of the subject by himself and his rating by his twin or sib; between his self rating and his rating by a friend; nor between his rating by twin or sib and by his friend. However, with very few exceptions the twins were rated more nearly alike by their friends than were the sibs. Perhaps the friend's rating was influenced somewhat by the fact that he was rating twins.

As shown by the intraclass correlations twins tend to be less alike on personality traits than on either intellectual or physical traits. This finding is in keeping with most other investigations of twins involving a study of personality traits. On the basis of the tests used the correlations indicate less agreement for twins than for ordinary sibs on the measures of self-sufficiency, introversion-extroversion, social adjustment, and masculinity-femininity. The difference is most marked (critical ratio 1.96) on the measure of social adjustment as determined by the Maller Character Sketches.

The mean percentile ratings tend to show that both the twins and the sibs approximate an average group, *i.e.* the average percentiles range from 45 to 55, on most traits measured. However, the twins as a group are more submissive, have less self-confidence and rate lower in mental health than do the sibs as indicated by lower mean percentile scores. Not only were the twins somewhat more submissive as a group than were the sibs, but the correlations indicate that a pair of twins may tend somewhat more to develop opposite trends in dominance and submission than do ordinary sibs. The difference, however, is not statistically significant.

The groups were too small to secure statistically reliable differences or to justify drawing any general conclusions. However, the findings tend to agree with studies involving larger groups of twins and suggest

that growing up as a twin may be a handicap in making a wholesome adjustment. The results suggest that were it possible to study a larger number of twins, twinning as a social factor might be proved significant in the development of personality. While the study may have little or no general significance, the analysis of personality based on the tests used helped in diagnosing a possible basis for the difficulty which the twins were encountering in making satisfactory social adjustments.

BIBLIOGRAPHY

1. Carter, H. D.: "Twin similarities in occupational interests." *Journal Educational Psychology*, Vol. XXIII, 1932, pp. 641–655.
2. Carter, H. D.: "Twin similarities in personality traits." *Journal Genetic Psychology*, Vol. XLIII, 1933, pp. 312–321.
3. Newman, H. H., Freeman, F. N., and Holzinger, K. J.: *Twins: A Study of Heredity and Environment.* The University of Chicago Press, Chicago, 1937, pp. 369 (p. 35).
4. Portenier, Lillian: "Aspects of socio-economic status which differentiate university students." *Journal of Applied Psychology*, June, 1938.
5. Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh and London, 1925, pp. 239.

BOOK REVIEWS

MARIE SKODAK. *Children in Foster Homes; a Study of Mental Development.* Iowa City: University of Iowa Studies in Child Welfare, 1939, 16, No. 1, pp. 141.

To explain the deep controversial hue that has characterized the problem of the origin of differences in human ability would probably require a careful dissection of the whole ideological *milieu* in which we live. Many investigators have found it impossible to attack the problem as an issue of pure science, but have envisaged from the start the applied aspects which are closely tied up with their value-systems. There is, in fact, less of a cleavage between pure research and application than is true of many scientific fields, since the most crucial data for the nature-nurture problem are obtained not through laboratory experiments, but through experiments which society itself has undertaken for social goals. This circumstance seems to lower the emotional thresholds of research workers, but paradoxically it places upon them an added responsibility to disengage themselves from emotional biases. If the results of scientific studies are of the kind that find immediate application, errors are costly to society. On the other hand, sound data have immeasurable social value.

Studies of foster children lend themselves to the clarification of a variety of developmental problems which are really inaccessible to any other approach. The inclusion by the Iowa Station of Child Welfare, therefore, of a long-term project on foster children in its research program on factors underlying the development of intelligence, should insure a background of crucial data against which to check and evaluate the many related studies pursued by the Iowa group. Moreover, because the Iowa project, unlike previous studies of foster children, deals with children of preschool ages, it should make headway into a conspicuous gap in a mosaic research structure to which other research workers have for years been contributing.

The data assembled in Skodak's study, which brings the Iowa foster child project up to date in monograph form, appear adequate to accomplish the foregoing desiderata. The test results, to the degree that they are trustworthy on the very young Subjects of this investigation, give evidence for environmental effects during the preschool years which are probably as marked as those in later years. At the same time they give evidence for hereditary differentia which increase with age even within the narrow age-range studied.

But these inferences are not the ones which Skodak imputes to the results—to data which in themselves are probably of sufficient importance to constitute basic source material for years to come. She offers interpretations, conclusions, and social implications, whose referents cannot be found either in her own material or in any other with which the reviewer is familiar. Unmooring her bark from the preschool-age boundaries of her investigation, and ignoring the findings of her predecessors on older foster children, she writes, first for "science":

> Evidence from this study indicates that it is the home rather than the child's true-family background which for practical purposes sets the limits of his mental development. The generalization which springs as a corollary is that the rather high relationship reported between the intelligence of own parents and older own children is probably very largely the result of environmental impacts on the child, the environment being largely determined and governed by the parents.

But how to explain the distinctly higher correlations previously found in the school-age range between IQ of own children and home variables, including parental intelligence, than between IQ of foster children and the same variables in the foster home? How to explain, indeed, the higher correlation (.33) among Skodak's older preschool foster children with the educational status of their own parents, with whom they had never lived, than with that (.18) of the foster parents who took them in early infancy?

Skodak continues as a transition to the sweeping application she is about to make:

> Placing the child in a good foster home at the youngest possible age makes for development equal to that of own children in similar homes.

Actually the Skodak report contains no data upon own children in homes similar to the foster homes. But why does she overlook the published data (from Stanford and Minnesóta) which are available upon school-age children, which show own children in superior and inferior homes to deviate *appreciably more from the average* than do foster children placed during infancy in comparable homes?

Finally the deduction for the applied field, as a tocsin to the placement agencies:

> Since the relationship between true-family background and the child's mental development is approximately zero and since knowledge of the child's true-family history has no predictive value, the use of true-family histories as a basis for the placement of the child has little or no justification.

This statement, in its way, seems quite as incautious as the statements of the "maturationist" group—Gesell *et al.*—who advocate tests during the first year of life (now known to have little long-range predictive power) as a safe basis for fitting the child to the foster home. Skodak's own "preschool" data show the predictibility of certain data regarding the true parents—though not high for any one variable by itself—to be about as high as and sometimes higher than corresponding data regarding the foster parents with whom the children are living. Other data, including those upon identical twins reared apart, show that family resemblance in intelligence appears and persists on the whole, even when environment has no opportunity to contribute to this resemblance. In its "social implication" Skodak's espousal of a policy of placement without regard to family background represents a grave retrogression. It cannot be dismissed with a light comment—not when one reflects upon the years it took to develop reasonable precautions in placement work, and not when one is personally acquainted with the life-long disappointments and frustrations that follow in the train of poor placement work. Not less use but more use of family histories, and emphasis upon *fuller family histories*, is the implication for the applied field contained in every study of foster children known to the reviewer.

. . . .

If Skodak's conclusions and her counsel are incompatible with or far exceed the actual data, what, then, do the data show?

Two basic groups of foster children were used as Subjects—one group of one hundred fifty-four cases who had been placed in permanent homes (chiefly adoptive homes) before the age of six months; and another group of sixty-five cases placed between the ages of two and five and one-half years after living under conditions of environmental deprivation.

In the first group called the "0 to 6 months group," the following results seem significant:

(1) Mean IQ above average at all ages; an apparent (but probably not an actual) decline of IQ with age in the age range studied.

(2) Little or no differentiation according to true family background *or* socio-economic level of foster home in the youngest age-groups (six to eighteen months), but increasing differentiation with age in both variables.

(3) Selective placement shown by a correlation of .30 between the mid-parent education of the true parents and the foster parents; also

by tabulations which show slight but not consistently positive relationships between the occupational classification of the foster fathers, and such variables as were recorded in the true-family background—IQ of mother, education of mother, and occupation of father.

Skodak's inference that the mental development of the children has been accelerated through foster home influence depends in turn upon the assumption that the IQ level would have been lower if the children had been reared in an average environment. Data upon the standardization of the tests, and upon the probable intellectual level of the true parents, are relevant in this connection.

Goodenough's careful study of the Kuhlmann-Binet scale indicates that it probably yields IQ's that are five or six points too high in the age-range two to four. Moreover, Kuhlmann's original data suggest that the allowance for "leniency" should be several points greater for children under two than for those of two to four, and several points less for children of five or six. There are also data which indicate that the 1916 Stanford-Revision, like the Kuhlmann, gives scores that are a little high during the preschool years.

This would mean that the superiority of intelligence scores in Skodak's group is less than appears from the face-values of the IQ's, particularly at the younger ages where the standardization is most open to question. It probably means, also, that the apparent decrease of IQ with age in Skodak's group is an artifact, since it agrees closely in amount with the decrease occurring in the original Kuhlmann age norms. Instead of a mean IQ of one hundred twenty in infancy falling to one hundred eight by age six (Skodak's Table 5), it would thus seem probable that the mean IQ is about seven points above normal throughout the range. This is in general agreement with the findings in studies of school-age foster children.

From the data available upon the true parents of the foster children it seems safe to conclude that their mental endowment was not above average, but it is questionable whether it was as far below average as the author infers from her material. In calculating IQ's of the true mothers, sixteen years was used as the maximum divisor, yielding a mean of 87.7. Age sixteen, however, is known to give an "IQ" somewhat too low for adults. Until the new Stanford revision introduced correction factors for computing adult IQ's, some psychometricians even advocated the use of fourteen years as a maximum divisor, which if defensible would bring the mean IQ of the true mothers practically to one hundred. Data upon the schooling of the true parents furnish corroborative evidence; with a median of ten school

grades completed the average IQ level of mothers and of fathers could hardly be much below one hundred. If this reasoning is sound, the children of Skodak's group have about seven points of IQ increment attributable to environment—which is again in accord with the findings in studies of school-age foster children.

The fact that the *youngest* children show an IQ increment which is not dependent upon the economic or cultural variables identified in the homes suggests that the factors there responsible may be associated with the passion for childhood which causes a couple without children to adopt one or more.

Since an increase with age is found in IQ differentiation following both foster family and true family background—selective placement not being effective enough wholly to account for either type of differentiation in terms of the other—we may conclude that the nature of the "mental ability" and of the environmental impact changes as the children emerge from infancy. Though merging with capacities anchored in the biological organism, the "nurture effect" also becomes more clearly traceable to socio-economic factors. There appears to be a peak (though not a statistically reliable peak) at age four in the relationship of IQ to variables in the foster home. This age may possibly represent an actual turning-point in an inverted U-shaped curve, since we know from previous studies that school-age foster children show hierarchies with respect to foster home variables that are far less marked than with Skodak's older preschool foster children.

. . . .

A comment must be added regarding the analysis of a subgroup of sixteen children whose true mothers were diagnosed as feebleminded on the basis of tests and other criteria. The mean IQ of these mothers was reported as 66.4 (which would, however, approach eighty if age fourteen rather than age sixteen should be used as a maximum divisor). The mean education of the mothers was seven and six-tenths grades, and of the true fathers, nine grades, as compared with nine and nine-tenths and ten and two-tenths for the true parents as a whole.

Skodak points out that some selection was operative in that the children were placed in homes slightly below the average of the entire group. She does not remind us, although it is surely relevant, (cf. p. 8) that "All children included in this study were judged by examining physicians to be physically normal and in good health at the time of placement," and (p. 9) that all children for whom legal adoption was sought spent a year of probationary residence in the home.

It has been the experience of the reviewer that children showing gross mental defects are usually identified through such precautions, and that only the more promising children of poor true-family background reach adoptive status. It is of interest to note that this subgroup at the time of first test was about five months older than the entire group of foster children at time of first test, which would probably indicate a longer probationary period.

On first examination (at mean age 2–5.5) the subgroup averaged practically the same in IQ as the group as a whole (at 2–0.7). On second examination (at mean age 4–11.5) the subgroup averaged four points lower than the group as a whole (at mean age 4–3.6). When the eight children placed in homes of the three upper occupational classifications and the eight placed in homes of the four lower occupational classifications were compared with children of the entire group who were placed in foster homes of corresponding levels, differentiation of practically the same kind and amount was found. But since the first test (*i.e.* the pre-adoption test) was not given until mean age 3–4.5 to the children placed in homes of the upper groups, and was given at mean age 1–7.9 to the children placed in homes of the lower groups, there was clearly more opportunity for sifting out unpromising children from the homes of superior status.

The following points need to be cleared up before the data on the subgroup of children of mentally deficient mothers will permit any conclusions to be safely drawn:

Were precautions taken to identify defective children of defective parents, with consequent selection of children available for the present study as the cream of the crop?

Were precautions more rigorous in the case of children placed in homes of higher occupational status than in homes of lower occupational status?

Will the children of defective parentage, who tested normal or above during the preschool years, continue to do so?

. . . .

The second group of sixty-five children, called by Skodak the two-to-five-year group, are of chief interest for the data on mental tests administered before and after placement in foster homes.

The true-family background of these cases was considerably inferior to that of the 0–6 months group on which the main study was based, and the foster homes were not as superior as those in which the 0–6 months group were placed. In addition to the early influence of

inferior home background, all of the children had been exposed for varying periods to an orphanage environment known to be unfavorable to mental development.

The mean IQ of the entire group at the time of "last preplacement examination" was 98.5 at a mean age of 3–4.2. "This group," writes the author (p. 113), "represents a selection from the upper levels of the institution population, in that children who were judged to be mentally retarded were not placed in foster homes."

The mean age at placement was 3–7.6. Following a period of at least one year in foster homes the children were re-examined, now showing a mean IQ of 104.2. For twenty-four cases who had a second examination two years after placement, the mean before placement was 97.9, 102.0 at the first postplacement test, and 107.7 at the second. For twenty children who received two preplacement tests while residing at the orphanage, there was no systematic shift in IQ—hence the increase of IQ's during foster home care was quite reasonably judged to be environmental in source.

No correlation was found between the factors of educational and occupational status of the true parents and child's IQ, but as only six fathers and seven mothers were known to have education beyond the eighth grade, and as nearly all the fathers belonged to the semi-skilled or unskilled labor groups, and as, further, the IQ distribution of the children was curtailed by eliminating retarded children, correlations could scarcely be expected. Though only a straw, it is of interest, however, that the seven children whose mothers had attended school beyond the eighth grade gained ten and eight-tenths points in IQ after foster home placement, whereas the group of thirty-nine with mothers of eighth-grade education or less gained only five and four-tenths points, and the nineteen cases for whom maternal education was unknown gained only three and seven-tenths points. Data are also reported upon the relationship of IQ and of gains in IQ to the socio-economic factors in the foster homes. Selective placement (fairly marked in this group), however, renders these analyses difficult of interpretation.

. . . .

If, as seems probable from this investigation viewed in the total setting of developmental studies, the nature of abilities appraised by intelligence tests *changes* during the preschool age period, one may ask whether there is any permanent value in "framing" the preschool home

environment to produce maximum IQ's. In so far as the final ceiling of IQ is concerned,[1] one would probably answer in the negative, but related aspects of social and personality development cannot here be left out of consideration.

One of the most challenging of all studies to come out of the fertile research program at Iowa was that of Lois M. Jack, who demonstrated in preschool children the dependence of social traits (ascendance and submission) upon the acquisition of knowledge and skills. Clinical studies offer corroborative evidence, and suggest that the preschool ages constitute the period *par excellence* not only for developing constructive attitudes toward tasks, but for integrating these with a sense of personal value and with feelings of security in social relationships. To accomplish this integration, however, would not necessarily mean that maximal mental growth during the preschool years was necessary or even desirable for all children. Rather, it would seem that the growth of adaptive behavior should be stimulated and guided —often in special areas instead of in all areas—so as to achieve a harmonious balance with the expanding personality needs of each particular child.

BARBARA S. BURKS.
Carnegie Institution of Washington, Cold Spring Harbor.

C. C. BENNETT. *An Inquiry into the Genesis of Poor Reading.* New York: Bureau of Publications, Teachers College, Columbia University, 1938, pp. 139.

An attempt has been made to analyze the environmental background and the general life adjustments of retarded readers in order to discover causal factors involved in the origin of their disability.

[1] Although most of the Iowa studies on the effect of environment deal with preschool children, Wellman has presented a study of high-school pupils which has been interpreted by her as demonstrating permanent effects of preschool stimulation, and of school stimulation in general upon the IQ. The selection of Subjects in her several groups, however, appears to introduce complicating factors that would preclude such a deduction without further evidence. Although the children of her experimental group and her control group were matched for initial IQ's during the preschool years, these IQ's are not highly prognostic for individual cases. Consequently there was ample opportunity during the twelve years of the children's school life for the "matched" preschool IQ's to "regress" toward the means of the IQ distributions they represented, giving differences which favored the groups which had preschool training, and training in the progressive schools of the University of Iowa. It is not uncommon to find an average IQ level of 120 or over in private schools, which can be accounted for as readily through family stock as through school stimulation.

Reading is regarded essentially as an adjustment rather than as a skill. Subjects were taken from grades 2B, 3A and 3B. Fifty children retarded one year on the average were paired with respect to age and mental ability with fifty children making normal progress in reading. The experimental (retarded) group was compared with the control group on the basis of statements or ratings by teachers, parents, and the children themselves. No reliable differences between the two groups were found in occupational status of the father, home libraries, scholastic training of parents, coöperation between home and school, physical status, lateral dominance, social and personality adjustments, leisure interest, and verbal interests. An extensive review of related studies and the analysis of a series of case studies are included. The author rightly emphasizes the fact that certain factors which fail to differentiate the two groups may be important in the reading adjustment of a particular case.

The careful reader will find several points to criticise in this report: (1) Matching of the pairs was not accurate. There was over a year's difference in chronological age in several pairs. The same was true for MA. (2) Most of the data were highly subjective. Quantitative treatment of such material is no justification for inadequateness of the original data. (3) The author tends to magnify slight and statistically unstable differences, especially in summarizing. For example, a critical ratio of two is stated to be "suggestive of real divergence between the two groups." Most of the ratios were much smaller than the two.

The study might be evaluated by saying that elementary-school children retarded about one year in reading cannot be differentiated from those making normal progress in terms of subjective ratings by teachers, parents and the children themselves on home, school and personality adjustments. The author was stimulated to state a number of interesting hypotheses but further research must test them out.

MILES A. TINKER.

University of Minnesota.

PUBLICATIONS RECEIVED

DONALD C. AGNEW. *The Effect of Various Amounts of Phonetic Training on Primary Reading.* Durham, N. C.: Duke University Press, 1939, pp. 50. (paper.)

N. ELDRED BINGHAM. *Teaching Nutrition in Biology Classes.* New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 117.

CHARLOTTE BÜHLER. *The Child and His Family.* Translated by Henry Beaumont. New York: Harper and Brothers, 1939, pp. 187.

OLIVER M. BUTTERFIELD. *Love Problems of Adolescence.* New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 212.

JOHN R. CLARK, ARTHUR S. OTIS, and CAROLINE HATTON. *Primary Arithmetic through Experience.* Yonkers, N. Y.: World Book Co., 1939, pp. 217.

LAWRENCE E. COLE. *General Psychology.* New York: McGraw-Hill Book Co., 1939, pp. 688.

EDMUND S. CONKLIN and FRANK S. FREEMAN. *Introductory Psychology for Students of Education.* New York: Henry Holt & Co., 1939, pp. 557.

EDITH A. DAVIS, and ESTHER MCGINNIS. *Parent Education.* Minneapolis: The University of Minnesota Press, 1939, pp. 153.

LESTER DIX. *A Charter for Progressive Education.* New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 107.

JACK W. DUNLAP. *Work in Statistical Method with Special Reference to the Social Sciences.* New York: Prentice-Hall, Inc., 1939, pp. 145. (paper.)

Education for Citizenship in Elementary Schools. The Association for Education in Citizenship. London: Oxford University Press, 1939, pp. 312.

JOHN WHARTON FAY. *American Psychology before William James.* New Brunswick, N. J.: Rutgers University Press, 1939, pp. 240.

ELLIS FREEMAN. *Principles of General Psychology.* New York: Henry Holt and Company, 1939, pp. 530.

FRANK N. FREEMAN. *Mental Tests, Their History, Principles and Applications.* Boston: Houghton Mifflin Co., 1939, pp. 460.

BRUNO FURST. *Use Your Head!* New York: Funk and Wagnalls Co., 1939, pp. 289.

CHESTER ROY GARVEY. *The Activity of Young Children during Sleep.* Minneapolis: The University of Minnesota Press, 1939, pp. 102.

A. R. GILLILAND and E. L. CLARK. *Psychology of Individual Differences.* New York: Prentice-Hall, Inc., 1939, pp. 535.

CLARENCE T. GRAY and DAVID F. VOTAW. *Statistics Applied to Education and Psychology.* New York: The Ronald Press Co., 1939, pp. 278.

COLEMAN R. GRIFFITH. *Psychology Applied to Teaching and Learning.* New York: Farrar & Rinehart, Inc., 1939, pp. 650.

J. P. GUILFORD. *General Psychology.* New York: D. Van Nostrand Co., Inc., 1939, pp. 630.

HENRY HANSBURG. *An Experimental Study of the Effect of the Use of the Print Shop in the Improvement of Spelling, Reading, and Visual Perception.* New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 84.

HELEN HEFFERNAN and OTHERS. *Music Education in the Elementary School.* Sacramento: California State Department of Education, 1939, pp. 152.

HARRY WALKER HEPNER. *It's Nice to Know People Like You.* New York: D. Appleton-Century Co., 1939, pp. 172.

KARL J. HOLZINGER and FRANCES SWINEFORD. *A Study in Factor Analysis: the Stability of a Bi-factor Solution.* Chicago: The University of Chicago, 1939, pp. 91. (paper.)

WILLIAM JAMES. *Talks to Teachers on Psychology.* New Edition. Introduction by John Dewey and William H. Kilpatrick. New York: Henry Holt and Co., 1939, pp. 238.

ARTHUR T. JERSILD and MARY D. FITE. *The Influence of Nursery School Experience on Children's Social Adjustments.* New York: Bureau of Publications, Teachers Coll., Columbia Univ., 1939, pp. 112.

MARGUERITE WILKER JOHNSON. *Verbal Influences on Children's Behavior.* Ann Arbor: University of Michigan Press, 1939, pp. 191.

CHARLES H. JUDD. *Educational Psychology.* Boston: Houghton Mifflin Co., 1939, pp. 566.

GLENN M. KENDALL. *Organization and Teaching of Social and Economic Studies in Correctional Institutions.* New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 159.

R. S. KIFER and H. L. STEWART. *Farming Hazards in the Drought Area.* Works Progress Administration, Research Monograph XVI. Washington: Government Printing Office, 1938, pp. 219. (paper.)

MARYELLEN MAHER LOMBARDI. *The Inter-trait Rating Technique.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1938, pp. 99.

VIKTOR LOWENFELD. *The Nature of Creative Activity.* New York: Harcourt, Brace & Co., 1939, pp. 272.

FREDERICK H. LUND. *Emotions.* New York: The Ronald Press Co., 1939, pp. 305.

A. M. MACMEEKEN. *The Intelligence of a Representative Group of Scottish Children.* London: University of London Press, 1939, pp. 144.

A. R. MANGUS. *Changing Aspects of Rural Relief.* Works Progress Administration, Research Monograph XIV. Washington: Government Printing Office, 1938, pp. 238. (paper.)

WILLIAM A. MCCALL. *Measurement.* New York: The Macmillan Co., 1939, pp. 535.

WILLIAM HENRY MIKESELL. *Mental Hygiene.* New York: Prentice-Hall, Inc., 1939, pp. 456.

PAUL MONROE, *Ed. Conference on Examinations,* 1938. New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 330.

WINONA L. MORGAN. *The Family Meets the Depression.* Minneapolis: The University of Minnesota Press, 1939, pp. 126.

EVELYN L. MUDGE. *Transfer of Training in Chemistry.* Baltimore: Johns Hopkins Press, 1939, pp. 76. (paper.)

JAMES L. MURSELL. *Educational Psychology.* New York: W. W. Norton & Co., 1939, pp. 324.

VICTOR H. NOLL. *The Teaching of Science in Elementary and Secondary Schools.* New York: Longmans, Green & Co., 1939, pp. 238.

W. O. D. PIERCE. *Air-war.* New York: Modern Age Books Inc., 1939, pp. 224. (paper.)

CARROLL C. PRATT. *The Logic of Modern Psychology.* New York: The Macmillan Co., 1939, pp. 185.

HOMER B. REED. *Psychology and Teaching of Secondary-school Subjects.* New York: Prentice-Hall, Inc., 1939, pp. 684.

FRANK S. SALISBURY. *Human Development and Learning.* New York: McGraw-Hill Book Co., 1939, pp. 513.

PETER SANDIFORD. *Foundations of Educational Psychology.* New York: Longmans, Green & Co., 1939, pp. 464.

Scottish Primary School Organization: Age-grade Classification, Time Allocation to Subjects. London: University of London Press, 1939, pp. 82. (paper.)

PAYSON SMITH, FRANK W. WRIGHT, and OTHERS. *Education in the Forty-eight States.* Staff Study Number One, The Advisory Committee on Education. Washington: Government Printing Office, 1939, pp. 198. (paper.)

GEORGE T. STAFFORD. *Sports for the Handicapped.* New York: Prentice-Hall, Inc., 1939, pp. 302.

Studies in Arithmetic. Reports on Investigations Relating to Present Practice and Teaching Methods in the Primary School. London: University of London Press, 1939, pp. 144.

GODFREY H. THOMSON. *The Factorial Analysis of Human Ability.* Boston: Houghton Mifflin Co., 1939, pp. 326.

H. A. TANSER. *The Settlement of Negroes in Kent County, Ontario.* Chatham, Ontario, Canada: The Shepherd Pub. Co., 1939, pp. 187.

ERNEST W. TIEGS. *Tests and Measurements in the Improvement of Learning.* Boston: Houghton Mifflin Co., 1939, pp. 490.

ALAN E. TRELOAR. *Elements of Statistical Reasoning.* New York: John Wiley & Sons, Inc., 1939, pp. 261.

M. J. A. VAN DER LUGT. *Un Profil Psychomoteur.* Paris: Aubier, 1939, pp. 126. (paper.)

J. E. WALLACE WALLIN. *Minor Mental Maladjustments in Normal People.* Durham, N. C.: Duke University Press, 1939, pp. 298.

GERTRUDE WALSH. *Sing Your Way to Better Speech.* New York: E. P. Dutton and Co., Inc., 1939, pp. 209.

GARNETTE WATTERS and S. A. COURTIS. *A Picture Dictionary for Children.* New York: Grosset and Dunlap, 1939, pp. 478.

L. W. WEBB and ANNA MARKT SHOTWELL. *Testing in the Elementary Schools.* Revised Edition. New York: Farrar and Rinehart, Inc., 1939, pp. 407.

E. C. WEBSTER. *Guidance for the High-school Pupil.* Montreal: McGill University, 1939, pp. 153.

GEORGE ROSS WELLS. *The Art of Being a Person.* New York: D. Appleton-Century Co., 1939, pp. 300.

LORIN F. WHEELWRIGHT. *An Experimental Study of the Perceptibility and Spacing of Music Symbols.* New York: Bureau of Publications, Teachers College, Columbia University, 1939, pp. 146.

E. H. WILLIAMSON. *How to Counsel Students.* New York: McGraw-Hill Book Co., 1939, pp. 562.

PAUL A. WITTY and CHARLES E. SKINNER, *Editors.* *Mental Hygiene in Modern Education.* New York: Farrar and Rinehart, Inc., 1939, pp. 539.

WALLER WYNNE. *Five Years of Rural Relief.* Works Progress Administration, Division of Social Research. Washington: Government Printing Office, 1938, pp. 160. (paper.)

SISTER MARY AGNES YEAKEL. *The Nineteenth Century Educational Contribution of the Sisters of Charity of Saint Vincent de Paul in Virginia.* Baltimore: Johns Hopkins Press, 1939, pp. 115. (paper.)

PAULINE V. YOUNG and CALVIN F. SCHMID. *Scientific Social Surveys and Research.* New York: Prentice-Hall, Inc., 1939, pp. 619.